



UNIVERSIDAD NACIONAL DE COLOMBIA

El modelo Lee-Carter para estimar y pronosticar mortalidad: Una aplicación para Colombia

Carlos Andrés Ochoa Molina

Universidad Nacional de Colombia

Facultad de Ciencias

Escuela de Estadística

Medellín, Colombia

2015

El modelo Lee-Carter para estimar y pronosticar mortalidad: Una aplicación para Colombia

Por:

Carlos Andrés Ochoa Molina

caochoam@unal.edu.co - ochoacarlosandres@gmail.com

Tesis o trabajo de grado presentada como requisito parcial para optar al título de:
Magíster en Ciencias-Estadística

Director: Norman Diego Giraldo Gómez, M.Sc.
Profesor Asociado, Escuela de Estadística, Facultad de Ciencias
Universidad Nacional de Colombia Sede Medellín

Línea de investigación: Actuaría

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística, Maestría en Ciencias-Estadística
Medellín, Colombia

2015

Declaración

Me permito afirmar que he realizado el presente trabajo de grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en la propia tesis. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los he reconocido en el presente trabajo. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de tesis o publicación.

Medellín, Antioquia, Mayo 22 de 2015

(Carlos Andrés Ochoa)

Agradecimientos

A la Universidad Nacional de Colombia y a la Escuela de Estadística de la Facultad de Ciencias por contribuir al fortalecimiento de mi perfil académico y profesional. Especiales agradecimientos al Profesor Norman Giraldo por su apoyo y contribución al presente trabajo. Además, se agradece a la Oficina de Planeación Académica de la Universidad Nacional de Colombia, sede Medellín por proporcionar los datos para la estimación de la población universitaria.

Resumen

El modelo de dinámica de mortalidad de Lee-Carter ha motivado un cambio en el análisis de la evolución poblacional apoyando nuevas hipótesis, formulaciones y herramientas para generar conclusiones al respecto. En este estudio se presentan dos aplicaciones del modelo Lee-Carter. En principio se ilustran tres herramientas para generar su estimación a través de descomposición singular y modelos lineales generalizados. Partiendo de la conclusión de que la metodología del modelo log bilineal de Poisson presentado por [Wilmoth, 1993] y [Brouhns et al., 2002], estimado con la librería `gnm` de **R** elaborada por [Turner & Firth, 2015] genera el mejor ajuste, se realizan dos análisis: i) Se estima un modelo para la dinámica de mortalidad en Colombia para los años 1951 – 1999 y se generan proyecciones hasta el año 2011, ii) Se analiza la dinámica de la mortalidad académica modalidad pregrado en la Universidad Nacional de Colombia sede Medellín en los periodos semestrales entre el 1989 -02 y 2006-01 y, de forma similar, se realizan las proyecciones correspondientes para 10 periodos semestrales entre el 2006-02 y el 2011-01. Los resultados sugieren un decaimiento significativo en la mortalidad poblacional en Colombia para Hombres y Mujeres, lo cual sugiere y motiva un análisis desde el punto de vista previsional para el sistema de seguros y pensiones. En el caso académico, las conclusiones señalan una mayor deserción estudiantil en los semestres iniciales y una disminución en los semestres necesarios para lograr el grado. En general, el modelo Lee-Carter prueba ser una herramienta útil para concluir sobre los movimientos en el tiempo de la masa poblacional relacionados con diferentes tipos de mortalidad.

Palabras clave: Descomposición singular; Modelos lineales generalizados; Lee-Carter; Pronósticos; Tasas de Mortalidad.

Abstract

The Lee-Carter methodology for modelling the dynamics of mortality rates has motivated a change in the way evolution of populations is analysed supporting new hypothesis, formulations and tools to generate conclusions. This study shows two applications of the Lee-Carter framework. First, we elaborate on three tools to estimate the model parameters through singular decomposition and linear models. Building on the conclusion that the methodology of the log bilinear Poisson model presented by [Wilmoth, 1993] and [Brouhns et al., 2002], which is estimated using the `gnm` library in **R** developed by [Turner & Firth, 2015] generates the best fit, we generate two analysis: i) The first analysis presents a model for the dynamics of mortality rates in Colombia for years 1951 – 1999, followed by forecasts up to 2011, ii) The second analysis shows the dynamics of academic mortality in undergraduate students of Universidad Nacional de Colombia, Medellin's branch for the academic terms (semesters) between 1989-02 and 2006-01, and in a similar way the corresponding projections for ten terms from 2006-02 to 2011-01. Evidence suggests a significant decay in mortality rates for Colombia's population for both genders, which encourages and motivates a grater analysis from the insurance and retirement system perspective. In the academic case the evidence shows a greater desertion in students of first academic terms and a decrease in the number of terms to graduate. In general terms the Lee-Carter model proves itself as a useful tool to conclude about the dynamics of populations through time, related to different types of mortality.

Keywords: Singular value decomposition; Generalized linear models; Lee-Carter; Forecasts; Mortality rates.

Contenido

Declaración	IV
Agradecimientos	V
Resumen	VII
Abstract	VIII
Índice de figuras	XI
Índice de tablas	XII
1. Introducción - Planteamiento del problema	1
2. Modelo Lee-Carter	3
2.1. Definiciones	4
2.2. Las funciones de la tabla de vida	5
2.3. El modelo	7
2.3.1. Matriz de descomposición singular (SVD)	8
2.3.2. Modelos lineales generalizados	10
3. Herramientas para la estimación del modelo Lee-Carter	12
3.1. Librería Demography	12
3.2. Librería ILC	13
3.3. Librería gnm	13
4. Estimación de la mortalidad en Colombia	15
4.1. Comparación entre modelos	15
4.1.1. Modelo LC ajustado a datos de hombres	15
4.1.2. Modelo LC ajustado a datos de mujeres	20
4.2. Estimación del modelo a través de gnm	26
4.3. Proyección de mortalidad	28
5. Estimación de la población universitaria	33
5.1. Estimación del modelo a través de gnm	36

5.2. Proyección de mortalidad	40
6. Conclusiones y recomendaciones	44
6.1. Conclusiones	44
6.2. Recomendaciones	45
Bibliografía	46

Lista de Figuras

4-1.	Comparación entre parámetros α_x , β_x y κ_t para los modelos Demography, ilc y gnm para hombres. Para cada uno de los parámetros α_x , β_x y κ_t se presenta normalizado restando su respectiva media en aras de poder hacer comparaciones ya que los valores generados con la librería gnm presentan diferencias debido a su algoritmo de estimación	17
4-2.	$\mu_{x,t}$ estimado para hombres con librería Demography	18
4-3.	$\mu_{x,t}$ estimado para hombres con librería ilc	19
4-4.	$\mu_{x,t}$ estimado para hombres con librería gnm	20
4-5.	Comparación entre parámetros α_x , β_x y κ_t para los modelos Demography, ilc y gnm para mujeres. Para cada uno de los parámetros α_x , β_x y κ_t se presenta normalizado restando su respectiva media en aras de poder hacer comparaciones ya que los valores generados con la librería gnm presentan diferencias debido a su algoritmo de estimación	21
4-6.	$\mu_{x,t}$ estimado para mujeres con librería Demography	22
4-7.	$\mu_{x,t}$ estimado para mujeres con librería ilc	23
4-8.	$\mu_{x,t}$ estimado para mujeres con librería gnm	24
4-9.	Residuales Deviance para el modelo gnm por edad y año en hombres y mujeres	26
4-10.	κ_t estimado (1951-1999) y proyectado (2003-2011)	30
4-11.	$\log(\mu_{x,t})$ estimado (1951-1999) y proyectado (2003-2011) - Hombres	31
4-12.	$\log(\mu_{x,t})$ estimado (1951-1999) y proyectado (2003-2011) - Mujeres	32
5-1.	Evolución de los semestres: 1989/02 - 2006/01	35
5-2.	Modelo LC para mortalidad académica, parámetros α_x , β_x , κ_τ y $\log(\mu_{x,\tau})$ estimados por Max.Ver.	36
5-3.	$\log(\mu_{x,t})$ para los 17 semestres a través de las 34 cohortes analizadas	38
5-4.	Residuos (deviation) del Modelo LC	39
5-5.	Ajuste de κ_t con ARIMA y Loess, y pronóstico a 10 semestres	41
5-6.	Pronósticos de 17 semestres a partir del 2011-01 con modelo LC usando Loess y ARIMA, datos observados para 8 semestres (Fuente: Oficina de Planeación, Universidad Nacional de Colombia Sede Medellín)	43

Lista de Tablas

4-1. Datos de expuestos al riesgo y muertes referentes al año 1993 para ambos generos ($E_{x,t}$ y $d_{x,t}$)	16
4-2. Residuales tipo Deviance de la estimación de $\hat{d}_{x,t}$ discriminados por género .	25
4-3. Residuales tipo Pearson de la estimación de $\hat{d}_{x,t}$ discriminados por género . .	25
4-4. Valores de $\hat{\alpha}_x$ y $\hat{\beta}_x$ a través de la librería gnm	28
4-5. Valores de $\hat{\kappa}_t$ a través de la librería gnm	28
5-1. Comparación de residuales Deviance y Pearson	40
5-2. Semestre 2011-01, pronosticada con Loess y Arima; Fuente de datos observa- dos: Oficina de planeación académica, Universidad Nacional Sede Medellín .	42

1 Introducción - Planteamiento del problema

La dinámica poblacional repercute directamente en la estabilidad financiera mundial. Particularmente, en la region latinoamericana se debe prestar especial atención debido a la estabilización de recientes reformas pensionales, las cuales apenas se están adaptando a la dinámica poblacional de cada país.

Hoy en día existe una gran cantidad de estudios que demuestran la dinámica poblacional en cuanto al crecimiento de la esperanza de vida, como dato concreto está la nueva tabla de mortalidad de rentistas aplicada por la legislación colombiana, a la cual las diferentes aseguradoras, fiducias y fondos de pensiones se tienen que ajustar. Por lo tanto, la pregunta que deberíamos plantear es si dichas tablas agrupan un fenómeno demográfico como lo es la mortalidad. En la mayoría de los casos, dichos ajustes llevarán a mayores pasivos o a menores mesadas pensionales debido a la mayor proyección de la esperanza de vida. Bajo el anterior escenario es válido preguntarnos si las entidades de previsión social están teniendo en cuenta el riesgo latente que planteamos aquí.

Como motivación al estudio de la dinámica poblacional se presentan algunos datos tomados del libro [Mundial, 1994]:

- El numero de personas mayores de 60 años pasará de 500 millones de personas a 1,400 millones para el año 2030.
- Por cuenta de los avances en medicina y la disminución de la fecundidad la población de los países en desarrollo envejecerá más rápido que la de los demás. En Bélgica se requerirá de 100 años para que la proporción de la población de más de 60 años pase del 9 % al 18 %, por el contrario en China este efecto tomará 34 años y 22 años en Venezuela.
- Las poblaciones de mayor edad requieren de mayor atención médica, lo cual implica unos costos financieros altos para la economía de un país. Cabe también mencionar que no es solo la atención médica sino el costo de sostenimiento en general.

En Colombia, bajo los diferentes regímenes pensionales, se trata de lograr una cobertura general para que la población logre tener una pensión; sin embargo, factores como el desem-

pleo y la informalidad contribuyen a que la mayoría de la población sea vulnerable a este riesgo. Este riesgo es de alto impacto económico y social en un país o región y por lo tanto debe ser ampliamente estudiado y analizado en aras de prever la materialización del riesgo y administrarlo de manera que no afecte la calidad de vida y el crecimiento económico del país.

En este estudio se incluye un capítulo en el que se modela la dinámica de mortalidad académica en los datos poblacionales (o de estudiantes) en la Universidad Nacional de Colombia Sede Medellín. Lo anterior se hace con el objetivo de ilustrar una aplicación adicional a un modelo de sobrevivencia actuarial. La aplicación de este tipo de modelos a mortalidad *académica* es válido en el contexto de ilustrar la manera como los estudiantes reaccionan a políticas institucionales o académicas en favor o en contra de su permanencia en la universidad, lo anterior se puede ver en dos sentidos; en mortalidad académica en el sentido del estudiante que renuncia a la universidad antes de culminar sus estudios, o en el sentido del estudiante que permanece en la universidad por más semestres de los que debe estar. En ambos sentidos es válido revisar como evoluciona la permanencia de un estudiante en la universidad a través de un método académico.

Este trabajo se desarrolla de la siguiente manera: El segundo capítulo ilustra el objetivo del modelo Lee-Carter (LC) el cual es acompañado de unas definiciones prácticas y el estado del arte de la modelación de la dinámica de mortalidad, además ilustra el proceso de estimación de parámetros a través de la metodología de descomposición singular y de modelos lineales generalizados. El capítulo 3 se enfoca en ilustrar las tres principales herramientas que se utilizan para realizar la estimación del modelo LC bajo las metodologías planteadas en el segundo capítulo usando R. El cuarto capítulo ilustra la aplicación del modelo LC a la información poblacional proveída por el DANE [Gómez, 1998] para los años entre el 1951 y el 1999 para hombres y mujeres. En este capítulo se compara la estimación del modelo bajo descomposición singular y bajo modelos lineales generalizados y no lineales generalizados. El capítulo termina por concluir acerca de la dinámica de la mortalidad en Colombia y realiza una proyección de esta mortalidad para el año 2011. En el capítulo 5 se realiza el ejercicio de análisis de la dinámica de mortalidad académica para la población estudiantil de la Universidad Nacional. En este capítulo se hace la estimación de parámetros del modelo y se realiza la proyección hasta la cohorte de 2011 – 01 terminando por concluir sobre esta. El estudio termina con el capítulo sexto anotando unas conclusiones y recomendaciones sobre el tema.

2 Modelo Lee-Carter

El objetivo de comprender la dinámica de la mortalidad, bajo la perspectiva demográfica, se basaba únicamente en estudios lineales a partir de la edad y la información conocida, lo cual generaba sobre-estimaciones e información inexacta. Lo anterior no permitía un análisis muy profundo en términos de actuaría de seguros y pensiones. Los estudios de este tipo iniciaron con modelos determinísticos como el de Gompertz en 1825 [Gompertz, 1825] el cual presenta una estimación satisfactoria de la mortalidad, sin embargo, sobreestima el indicador para edades superiores a 80 años, ver [Koissi & Shapiro, 2008]. El modelo de Heligman y Pollard (1980) [Helligman & Pollard, 1980] continuó avanzando en el tema, para lo cual, a partir de interpolaciones, dio pasos iniciales hacia el objetivo de proyección de mortalidad en aras de estructurar los estudios actuariales que sustentan los negocios de seguros y pensiones [Betzen, 2010]. En los años recientes las tendencias se han inclinado hacia el modelo estocástico presentado por Lee y Carter [Lee & Carter, 1992]. Desde entonces, este modelo ha sido utilizado para aplicaciones demográficas y actuariales en todo el mundo.

El modelo LC concebido para pronosticar mortalidad y analizar su dinámica fue definido en 1992, su intención más allá de analizar la interacción entre variables como tasa de mortalidad, esperanza de vida e índice de mortalidad se limita por los patrones existentes en esta temática, es decir, no incluye información relativa a la accidentalidad, avances en medicina, guerras u otros eventos que pueden marcar un punto de inflexión. El modelo basa su aproximación en la proyección de la tendencia histórica presentada por las variables, así mismo, su composición probabilística permite, a través de series de tiempo, generar análisis sobre el comportamiento futuro como lo pueden ser pronósticos e intervalos de confianza. A pesar de sus limitantes, la aproximación definida por Lee y Carter en este sentido es utilizada ampliamente en el medio demográfico y logra estar presente en un gran porcentaje de las investigaciones sobre el tema.

La diferenciación más importante presentada por Lee y Carter en su artículo fue la incorporación de la información en dos dimensiones, es decir, mortalidad a través de periodos de tiempo. Concretamente, el modelo asume que la dinámica de mortalidad responde a un parámetro generado por la regresión: el índice de mortalidad. A partir de este índice se puede proyectar el comportamiento de la mortalidad usando un modelo clásico de series de tiempo como lo es Box - Jenkins, ver [Lee & Carter, 1992]. De forma similar el modelo permite definir conclusiones sobre la esperanza de vida y tablas de mortalidad.

2.1. Definiciones

- **Tasas de mortalidad para una edad específica (Hazard rate)**

Define el número de muertes en un rango de edad específico para un tiempo determinado.

- **Tasa de mortalidad central**

Tasa de mortalidad calculada a partir de la información poblacional de mitad del año.

- **Tablas de mortalidad**

Son tablas de valores numéricos que exhiben variables relativas a la mortalidad y expectativa de vida de una cohorte o población por periodo. Entre las variables observadas en una tabla de mortalidad se encuentran: probabilidad de sobrevivencia, probabilidad de muerte, número de muertes entre edades específicas, promedio de personas vivas entre edades exactas, población total con una edad específica y la esperanza de vida para una edad específica.

- **Expectativa de vida**

Define la cantidad de años que vivirá una persona de una edad determinada si las condiciones de mortalidad se mantienen constantes en el tiempo.

- **Forma del perfil promedio (Age specific mortality)**

Describe la fuerza de la variación de la mortalidad ante cambios en los grupos de edad o cohortes

- **Índice de mortalidad**

Parámetro dependiente del tiempo, el cual identifica la dinámica creciente o decreciente de la mortalidad

- **Modelo Box - Jenkins**

Modelo de series de tiempo (ARMA) basado en media móvil autorregresiva.

En el modelo a analizar la variable de interés es la fuerza de mortalidad $\mu_{x,t}$ asociada a un individuo de edad x en el año t , entre más grande $\mu_{x,t}$ mayor es la probabilidad de muerte de este individuo. Dicho modelo se plantea definiendo $\log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t$, donde $\alpha_x, \beta_x, \kappa_t$ son los parámetros a estimar. Lo planteado en la ecuación de LC es una homotopía, es decir, la fuerza de mortalidad en t_1 es similar a la fuerza de mortalidad en las demás cohortes; la diferencia consiste en el parámetro κ_t , que mide el efecto de la cohorte, modulado por β_x .

La forma de $\exp(\alpha_x)$ define el patrón de mortalidad, pero el efecto de la cohorte lo cambia. La modelación de la dinámica en la mortalidad, como objetivo planteado en este documento se pretende analizar a partir de la proyección a través del tiempo del índice de mortalidad (k_t), el modelo inicialmente planteado por Lee y Carter propone la estimación a través de un ARIMA, ver [Lee & Carter, 1992].

El modelo LC original fué modificado por [Brouhns et al., 2002] para reformularlo como un modelo log-bilineal de regresión Poisson, concretamente, un modelo no lineal generalizado de tipo Poisson, aunque se plantean varias modificaciones como utilizar distribuciones Binomial Negativas y Binomiales en lugar de la Poisson. La estimación se realiza de tres diferentes maneras:

1. *SVD* - Descomposición singular de matrices a través de la función `svd` de R o la librería `demography`
2. *GLM* - Modelos lineales generalizados a través de la librería `ILC`
3. *GNM* - Modelos no lineales generalizados a través de la librería `GNM`

2.2. Las funciones de la tabla de vida

Las siguientes definiciones relacionadas con las tablas de vida o tablas de mortalidad se presentan en aras de describir conceptualmente las variables involucradas en el modelo. Las siguientes se presentan en términos de dinámica de mortalidad poblacional en general, mas adelante se especificará una aplicación particular a la *mortalidad académica* en universidades.

Denotamos por $S(x) := 1 - F(x)$ la función de supervivencia. Si $x \geq 0$ consideremos la variable aleatoria $T(x) := X - x | X > x$ con distribución $G(t) := \mathbb{P}(T(x) \leq t)$ para $t \geq 0$. La función $G(t)$ se interpreta como la probabilidad de muerte antes de t años habiendo llegado a x años de edad. La notación actuarial para esta función es ${}_tq_x$. Igualmente ${}_tp_x := 1 - {}_tq_x$. Cuando $x = 0$ se tiene que $G(t) = F(t)$, y cuando $t = 1$ se escribe $q_x = {}_1q_x$, igual con p_x . La función de fuerza de mortalidad es la base para la definición de modelos paramétricos. Esta función se define como $\mu_x := \frac{1}{S(x)} \frac{dF(x)}{dx}$, y se cumple la identidad

$${}_tp_x = e^{\left(-\int_0^t \mu_{x+s} ds\right)} \quad (2-1)$$

Asumiendo una población inicial de l_0 , se define una variable aleatoria binomial, $\mathcal{L}(x)$ tal que $\mathcal{L}(x) \sim \text{Bin}(l_0, S(x))$. La población que llega con vida a una edad x se define como el valor esperado de $\mathcal{L}(x)$, $l_x := E(\mathcal{L}(x)) = l_0 S(x)$.

De igual manera, el número de personas que mueren entre la edad x y $x + 1$ es una variable aleatoria binomial, $\mathcal{D}(x) \sim \text{Bin}(l_0, S(x) - S(x + 1))$. Su valor esperado se denota por $d_x = l_0(S(x) - S(x + 1)) = l_x - l_{x+1}$. Por tanto, $d_x/l_x = (S(x) - S(x + 1))/S(x) = q_x$. De (2-1) sigue la identidad

$$l_{x+1} = l_x e^{(-\int_0^1 \mu_{x+s} ds)} \quad (2-2)$$

El supuesto de fuerza de mortalidad constante por tramos establece que

$$\mu_{x+s} = \mu_x, \quad 0 \leq s < 1. \quad (2-3)$$

por tanto, al asumirlo, $l_{x+1} = l_x e^{-\mu_x}$ y $q_x = 1 - e^{-\mu_x}$. La función l_x proporciona el total de personas que llegaron a la edad x ; pero el total de personas que llegaron a la edad x y murieron antes de cumplir años se define de una manera diferente, se indica por L_x y se puede aproximar por $L_x = l_x - d_x/2$. La función L_x se denomina total de expuestos al riesgo, o exposición, también denotada por E_x . La función $q_x = d_x/l_x$ se denomina tasa de mortalidad. La función $m_x = d_x/L_x$ se denomina tasa central de mortalidad. Se comprueba que aproximadamente, $m_x = \mu_{x+1/2}$.

El siguiente concepto que se debe clarificar es el de cohorte. Una cohorte da una especificación a un grupo en la población, y lo caracteriza de la siguiente manera:

1. El grupo consta de una población inicial de l_0 .
2. Cada miembro del grupo tiene una probabilidad q_x de muerte al iniciar la edad x .
3. Al grupo no ingresan otras personas fuera de los l_0 iniciales.
4. Los tiempos de supervivencia en cada grupo de análisis se asumen variables aleatorias independientes, lo que sugiere que no hay relación entre las causas de mortalidad entre las diferentes cohortes.

Con lo anterior en mente, $l_{x,t}$ será el total de personas que llega a la edad x correspondientes a la cohorte t . n_c será el número total de cohortes observadas de tal manera que $t = t_1, \dots, t_{n_c}$. Si ω_t corresponde a la última edad a la que llegó la cohorte t entonces $x = 1, \dots, \omega_t$

Continuando con el concepto de cohorte podemos reescribir la ecuación (2-2) de la siguiente manera:

$$l_{x+1,t} = l_{x,t} e^{(-\int_0^1 \mu_{x+s,t} ds)} \quad (2-4)$$

2.3. El modelo

El modelo inicialmente planteado por Lee y Carter parte de la relación lineal entre el logaritmo de la tasa central de mortalidad y las variables explicativas del fenómeno tales como edad y tiempo. Se debe anotar que este modelo es válido para poblaciones no estacionarias, en las cuales las funciones biométricas dependen de la cohorte t . El modelo es de la forma:

$$m_{x,t} = e^{\alpha_x + \beta_x \kappa_t + \epsilon_{x,t}} \quad (2-5)$$

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \quad (2-6)$$

donde $x = 1, \dots, \omega_t$, $t = t_1, \dots, t_{n_c}$ y $\epsilon_{x,t} \sim i.i.dN(0, \sigma^2)$. α_x , β_x y κ_t son los parametros a estimar.

En donde las variables se definen de la siguiente manera:

- μ_x : Fuerza de la mortalidad.
- $m_{x,t}$: Tasa de mortalidad central para la edad x en el tiempo t (Hazard rate). Número de muertes de edad x a $x + 1$ con respecto los que tenía x en la mitad del año t .
- $d_{x,t}$: Muertes registradas para la edad x en el periodo t .
- $E_{x,t}$: Expuestos al riesgo en la edad x correspondientes al periodo o cohorte t .
- α_x : Logaritmo del comportamiento medio de μ_x (Fuerza de la mortalidad).
- β_x : Descripción del patron de desviación del perfil de edad x a medida que k varía. Medida de variación de $m_{x,t}$ a medida que κ_t cambia.
- κ_t : Índice de mortalidad o tendencia en el tiempo de la fuerza de mortalidad. Este parámetro define la dirección de la esperanza de vida, ante una reducción en el índice de mortalidad habrá una mayor esperanza de vida. Este índice podrá ser negativo para así reflejar cambios en la dinámica de la tasa de mortalidad central.
- $\epsilon_{x,t}$: Término de error.

Debido al término bilineal $\beta_x \kappa_t$ estos parámetros del modelo no serán identificables (no podrán ser estimados) debido a que el modelo es invariante a transformaciones lineales, por ejemplo si α_x , β_x y κ_τ son una solución, entonces $\forall c \neq 0$, $\alpha_x - \beta_x c$, β_x y $\kappa_\tau + c$ también serán soluciones y de la misma manera, α_x , $\beta_x c$ y κ_τ / c serán soluciones; pero con las restricciones

$$\sum_{\tau} \kappa_{\tau} = 0, \sum_x \beta_x = 1 \quad (2-7)$$

el modelo queda identificable.

El modelo no se puede ajustar por regresión debido a que no tiene variables explicativas observadas en el término de la derecha de 2-6. Nótese (ver [Brouhns & Denuit, 2002] pag. 50) que se cumple (ignorando el efecto del error aleatorio)

$$\frac{\partial \log(m_{x,\tau})}{\partial \tau} = \beta_x \frac{\partial \kappa_\tau}{\partial \tau}$$

por lo que las edades x para las cuales β_x tenga valor absoluto más grande, son más sensibles a los cambios en la variación temporal de κ_τ .

Asumimos en este trabajo el supuesto de fuerza de mortalidad constante el cual garantiza uniformidad en la dinámica de mortalidad entre cada año, es decir

$$\mu_{x+s,\tau} = \mu_{x,\tau}, \quad 0 \leq s < 1. \quad (2-8)$$

por tanto

$$l_{x+1,\tau} = l_{x,\tau} e^{(-\mu_{x,\tau})} = l_{x,\tau} e^{-e^{\alpha_x + \beta_x \kappa_\tau}} \quad (2-9)$$

también se cumple

$$q_{x,\tau} = 1 - e^{-e^{\alpha_x + \beta_x \kappa_\tau}} \quad (2-10)$$

Este supuesto nos permite calcular la probabilidad $\mathbb{P}(T(x) \leq 1) = q_x$ como ${}_s q_x = 1 - e^{s \log(1-q_x)}$. Teniendo en mente el tipo de datos a usar para los modelos a continuación, en el que los datos se reportan por periodo (años para mortalidad poblacional, y semestres para mortalidad académica), se asume que la disminución poblacional se da al final de dicho periodo. Para analizar el impacto de este supuesto se requiere de datos poblacionales más detallados, lo cual es una imposibilidad técnica que nos lleva a asumir como válido el supuesto, ver [Bowers et al., 1986].

En la literatura actual se dan dos posibilidades de estimación de los parámetros del modelo. SVD (Descomposición singular) o a través de GLM (Modelos lineales generalizados).

2.3.1. Matriz de descomposición singular (SVD)

La estrategia original de estimación del modelo de LC (2-6) se basa en la minimización de la función

$$D(\underline{\alpha}, \underline{\beta}, \underline{\kappa}) = \sum_{x,t} (\log(m_{x,t}) - (\alpha_x + \beta_x \kappa_t))^2,$$

con respecto a $(\underline{\alpha}, \underline{\beta}, \underline{\kappa})$, $\underline{\alpha}, \underline{\beta} \in \mathbb{R}^{n_a}$, $\underline{\kappa} \in \mathbb{R}^{n_c}$. En forma matricial, si $M = [\log(m_{x,t})] \in \mathbb{R}^{n_a \times n_c}$, $\underline{1}_{n_c} = (1, \dots, 1)' \in \mathbb{R}^{n_c}$

$$D(\underline{\alpha}, \underline{\beta}, \underline{\kappa}) = \|M - \underline{\alpha}\underline{1}'_{n_c} - \underline{\beta}\underline{\kappa}'\|_2^2. \quad (2-11)$$

Donde $\|A\|_2 = \sqrt{\sum_i \sum_j |A_{i,j}|^2}$ es la norma Frobenius para matrices. A partir de la ecuación $\partial D / \partial \underline{\alpha} = \underline{0}$ las restricciones (2-7) permiten obtener como solución el vector dado por los promedios de las filas de M , $\hat{\underline{\alpha}} = \frac{1}{n_c} M \underline{1}_{n_c}$. Reemplazando esta solución en (2-11), el problema ahora consiste en minimizar con respecto a $(\underline{\beta}, \underline{\kappa})$ la función

$$D(\underline{\beta}, \underline{\kappa}) = \|M - \hat{\underline{\alpha}}\underline{1}'_{n_c} - \underline{\beta}\underline{\kappa}'\|_2^2. \quad (2-12)$$

La minimización en (2-12) es un caso particular del problema general de aproximar una matriz dada H por otra matriz de rango igual o menor, X , es decir, se trata de encontrar \hat{X} tal que

$$\hat{X} = \underset{rank(X) \leq r}{\operatorname{argmin}} \|H - X\|_2, \quad (2-13)$$

donde $r = rank(H)$. La solución \hat{X} se puede encontrar mediante la descomposición singular de la matriz H .

El teorema de descomposición singular de una matriz (ver [Shores, 2004], sec. 5.6) establece que: dada una matriz $H \in \mathbb{R}^{m \times n}$ se pueden definir dos matrices ortogonales (sus columnas son vectores ortogonales) $U \in \mathbb{R}^{m \times m}$, y $V \in \mathbb{R}^{n \times n}$, tales que $H = U \Sigma V'$, donde $\Sigma \in \mathbb{R}^{m \times n}$ es una matriz diagonal $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_p)$, con $p = \min(m, n)$ y $\sigma_1 \geq \dots \geq \sigma_p \geq 0$. Los valores σ_j se denominan valores singulares de H , las columnas de U los vectores singulares izquierdos y las de V vectores singulares derechos. Si $r \leq \min(m, n)$ es el rango de H se cumple $\sigma_r > 0$ y $\sigma_{r+1} = 0$. El siguiente resultado es un corolario de este teorema (ver [Shores, 2004], Cor.5.7,(6)). Si $U = [u_1, \dots, u_m]$, $V = [v_1, \dots, v_n]$ entonces se cumple la descomposición siguiente.

$$H = \sum_{j=1}^r \sigma_j u_j v_j'. \quad (2-14)$$

Si la matriz original es real, no se generarán números complejos ya que los valores singulares σ_j son reales no negativos y los vectores u_j, v_j son reales.

La solución \hat{X} al problema (2-13) para $rank(X) = h < r = \min(m, n)$ está dada por la siguiente expresión, establecida en [Eckart & Young, 1936]

$$\hat{X} = \sum_{j=1}^h \sigma_j u_j v_j'. \quad (2-15)$$

Aplicando este resultado a la matriz $H = M - \hat{\alpha} \underline{1}'_{n_c}$, asumiendo $n_a = m < n = n_c$, con $r = n_a$,

$$M - \hat{\alpha} \underline{1}'_{n_c} = \sum_{j=1}^h \sigma_j u_j v_j'. \quad (2-16)$$

En el trabajo original [Lee & Carter, 1992] utilizan una aproximación de rango $h = 1$ y por tanto, se estiman $\hat{\beta} = \sqrt{\sigma_1} u_1$ y $\hat{k} = \sqrt{\sigma_1} v_1$, respectivamente. Este resultado se puede colocar de manera informal como

$$M \approx \hat{\alpha} \underline{1}'_{n_c} + \hat{\beta} \hat{k}'. \quad (2-17)$$

En [Koissi & Shapiro, 2008] y [Brouhns et al., 2002] se analiza con detalle esta metodología.

2.3.2. Modelos lineales generalizados

El modelo original LC (2-6) presenta varias desventajas. Una de ellas es que asume homocedasticidad en los errores; pero este supuesto no se cumple con datos de tablas de vida, en los cuales, en las edades avanzadas al tener pocos datos para la exposición, los estimadores presentan alta varianza en contraste con los estimadores de edades tempranas, ya que $Var(\log(m_{x,t})) \approx 1/d_{x,t}$ (ver [Wilmoth, 1993], pag.2).

Para corregir este hecho [Brouhns et al., 2002] modificaron el modelo original, asumiendo $\mathcal{D}(x, t) \sim \text{Poisson}(E_{x,t} m_{x,t})$ en lugar de $\mathcal{D}(x) \sim \text{Bin}(l_0, S(x) - S(x+1))$, tal que

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t. \quad (2-18)$$

donde $d_{x,t}/E_{x,t} = m_{x,t}$; siendo $d_{x,t}$ las muertes de edad x para la cohorte t .

Una expresión equivalente del modelo es

$$\log(d_{x,t}) = \log(E_{x,t}) + \alpha_x + \beta_x \kappa_t. \quad (2-19)$$

El término $\log(E_{x,t})$ se denomina offset. Esta diferenciación se hace ya que el término observado es $d_{x,t}$ (muertes) y $E_{x,t}$ (expuestos al riesgo).

El método de máxima verosimilitud para estimación de la versión log-bilineal Poisson del modelo LC (2-19), está desarrollado en [Wilmoth, 1993] y [Brouhns et al., 2002] y complementado por [Renshaw & Haberman, 2009]. Como $\mathcal{D}(x, t) \sim \text{Poisson}(L_{x,t} m_{x,t})$ entonces la función log-verosimilitud basada en los datos $d_{x,t}$, $E_{x,t}$, y haciendo $\lambda_{x,t} = E_{x,t} m_{x,t}$, está dada por

$$l(d, \lambda) = \sum_{x,t} d_{x,t} \ln(\lambda_{x,t}) - \lambda_{x,t} - \ln(d_{x,t}!). \quad (2-20)$$

Se busca encontrar los valores en la matriz $\lambda = [\lambda_{x,t}]$ que maximicen (2-20). Como se cumple

$$\lambda_{x,t} = E_{x,t} m_{x,t} = E_{x,t} e^{\alpha_x + \beta_x \kappa_t}$$

entonces maximizar (2-20) con respecto a $\lambda_{x,t}$ equivale a maximizar

$$l(\alpha, \beta, \kappa) = \sum_{x,t} d_{x,t} \ln(E_{x,t} e^{\alpha_x + \beta_x \kappa_t}) - E_{x,t} e^{\alpha_x + \beta_x \kappa_t} \quad (2-21)$$

con respecto a $\alpha_x, \beta_x, \kappa_t, x = 1, \dots, n_a, t = t, \dots, tn_c$.

El estimador de máxima verosimilitud en el contexto de modelos lineales generalizados se obtiene minimizando la función deviance. Si $d_{x,t}$ es el conteo de bajas observado y $\hat{d}_{x,t} = E_{x,t} e^{\hat{\alpha}_x + \hat{\beta}_x \kappa_t}$, la función deviance se define como

$$D(d_{x,t}, \hat{d}_{x,t}) = 2 \sum_{x,t} \omega_{x,t} \left(d_{x,t} \log \left(\frac{d_{x,t}}{\hat{d}_{x,t}} \right) - (d_{x,t} - \hat{d}_{x,t}) \right), \quad (2-22)$$

donde $\omega_{x,t} = 0$ si $d_{x,t} = 0$ y $\omega_{x,t} = 1$ si $d_{x,t} > 0$. La minimización de la función de deviance se hace mediante el método de Newton-Raphson unidimensional, aplicado a cada parámetro $\theta = \alpha_x, \beta_x, \kappa_t$, el cual actualiza el valor de θ según la expresión

$$\theta_{j+1} = \theta_j - \frac{\frac{\partial D}{\partial \theta}}{\frac{\partial^2 D}{\partial \theta^2}} \quad (2-23)$$

3 Herramientas para la estimación del modelo Lee-Carter

3.1. Librería Demography

La librería Demography creada por Rob J Hyndman (2014), ver [Hyndman, 2015], permite realizar análisis de datos demográficos que incluyen el análisis de tablas de vida. El análisis generado por esta herramienta es basado en la metodología original planteada por Lee y Carter [Lee & Carter, 1992] de descomposición en valores singulares.

El siguiente código en R permite implementar este tipo de modelos:

```
modlc = lca(datos, interpolate = TRUE, adjust = 'none')
```

interpolate: Define una interpolación para tasas de cero mortalidad con base en datos de la misma cohorte de edades similares.

adjust: Define la metodología de ajuste para los coeficientes κ_t . Se usa "none" para evitar que tome el ajuste por defecto para la metodología Booth-Maindonald-Smith.

datos es un objeto creado con la función **demogdata** de la librería **demography**, con la instrucción

```
datos = demogdata(data=dx, pop=Ex, years=t, ages=x,  
type="mortalidad", label="Datos", name="cohortes")
```

data: Datos de muertes en la población.

pop: Valores poblacionales o expuestos al riesgo. En este sentido afecta el supuesto de mortalidad constante ya que el modelo asume que el registro es a 30 de Junio de cada año.

years: Vector que tiene los años de la muestra o cohortes respectivas.

ages: Las diferentes edades que se presentan entre la muestra.

3.2. Librería ILC

La estimación a través de máxima verosimilitud sobre el modelo lineal generalizado se realiza con la librería `Ilc` creada por [Butt & Haberman, 2009]. Inicialmente se debe crear un objeto que contenga la información de d_x, m_x, x, t de la siguiente manera:

```
datos = demogdata(data=dx, pop=Ex, years=t, ages=x,
type="mortality", label="Datos", name="cohortes")
```

A partir de esto se realiza la estimación con el código `lca.rh()` aplicado al objeto `datos`:

```
m1 = lca.rh(datos, mod='lc', interpolate=TRUE,
verbose = FALSE)
```

mod: A través de este parámetro se indica cuál de los modelos se va a ajustar. Particularmente se selecciona `'lc'` correspondiente al modelo LC. De esta misma manera se define la función `link` del modelo lineal generalizado.

interpolate: Igual a la librería `Demography`.

verbose: Define si en la ejecución del programa se imprime los valores actualizados del deviance.

3.3. Librería gnm

El modelo log-bilineal Poisson (2-19) se puede definir como un ejemplo de la clase de modelos no-lineales generalizados, definidos en: (ver [Davidian, 2009, Chapter 4]).

La librería `gnm` creada por [Turner & Firth, 2015] para estimación en modelos no-lineales generalizados, tiene como uno de sus ejemplos el modelo LC el cual se puede ver como un modelo no-lineal generalizado debido a la presencia del término bilineal. En [Currie, 2014] se analiza la aplicación de la función para estimar modelos lineales generalizados `glm()` para algunos de los modelos definidos en [Cairns et al., 2009], y la función `gnm()` para los modelos no-lineales generalizados. El objetivo de [Currie, 2014] es mostrar que estas funciones proveen una mayor flexibilidad frente a las posibilidades de las librerías `ilc` y `LifeMetrics`. En particular, estas funciones permiten estimar modelos con base en las tasas de mortalidad $q_{x,\tau}$ y utilizar distribuciones con efecto de sobredispersión. Lo anterior garantiza un mayor ajuste en el caso en que la varianza del modelo sea mayor a la esperada, lo cual se hace relevante en este caso por baja calidad en los datos observados, en específico con agrupación de edades. Aunque existen pruebas en **R** para analizar el efecto de sobredispersión como la llevada a cabo por la librería `AER` (`dispersiontest(md, trafo=1)`), cabe mencionar que dicho análisis

no hizo parte de la investigación de este trabajo por lo que se deja como recomendación para una próxima revisión. En [Turner & Firth, 2015] se encuentra una exposición muy detallada de las opciones de la función `gnm()` y su aplicación a la estimación del modelo LC.

El proceso de optimización llevado a cabo por la librería `gnm()` implementa el método de máxima verosimilitud en dos pasos: un paso utiliza una estimación mediante el método de Newton-Raphson, de manera iterativa para cada parámetro por separado, y un segundo paso actualiza todos los parámetros utilizando Newton-Raphson pero con matrices pseudo-inversas Moore-Penrose para evitar matrices singulares. En el modelo log-lineal Poisson la máxima verosimilitud es equivalente a la función deviance. Sin embargo, [Turner & Firth, 2015] aclaran que el algoritmo no produce siempre el mismo vector de parámetros. En cada corrida se produce uno diferente. Es posible que se deba a la implementación de valores iniciales aleatorios. Aun con esto, [Turner & Firth, 2015] especifican que el valor final de deviance permanece siempre invariante así como los valores ajustados. Adicionalmente, la función `gnm()` no utiliza las restricciones del modelo LC para garantizar identificabilidad. En [Currie, 2014], pag. 17, ec. (47,48,49) se establecen transformaciones para implementar estas restricciones al final de la estimación. En resumen, aunque se producen estimadores diferentes en cada corrida, no se considera que sea debido a que se alcanzó un mínimo local, ya que en ese caso la transformación mencionada produciría estimadores diferentes, lo cual sería un problema, el cual no sucede. Por tanto, se podría afirmar que al correr el programa con valores iniciales aleatorios y llegar después de transformaciones a un mismo vector de parámetros significa que, con valores de tolerancia y número máximo de iteraciones, sí se llega a un mínimo global.

La sintaxis para estimar un modelo consiste en definir las variables $\mathbf{t}, \mathbf{x}, \mathbf{dx} (d_{x,t}), \mathbf{Ex} (E_{x,t})$, contenidas en un `data.frame`, por ejemplo, `D`, con la instrucción

```
m1 = gnm(dx ~ x + Mult(x, t), offset = log(Ex),
family = "poisson", data = D)
```

formula (Ej. `dx ~ x + Mult(x, t)`): Fórmula en la cual se indica la relación entre los parámetros a estimar y el dato observado. En este caso será `dx` (Muertes) se regresará contra un término multiplicativo en t, x .

offset: Este parámetro indica que hay información adicional que debe incluirse en el análisis. En este caso serán los expuestos al riesgo.

family: Especifica la función link del modelo.

Esta opción permite utilizar otras familias de modelos exponenciales, como “quasi-poisson” con la cual se estima un modelo Poisson con parámetro de sobre-dispersión ϕ .

4 Estimación de la mortalidad en Colombia

En este capítulo se ilustra la estimación del modelo LC para cada una de las librerías mencionadas anteriormente con un enfoque en la librería en `gnm` en **R**. Este análisis incluye la revisión de las diferencias presentadas entre los hombres y mujeres en el periodo de estudio. Al finalizar el capítulo, a partir de la evidencia, principalmente en el análisis de residuales, recomendamos que la librería a utilizar para la proyección de la mortalidad es la `gnm` presentada por [Turner & Firth, 2015]. Así mismo, se incluye un análisis correspondiente a las proyecciones generadas, en las que se concluye a partir de la evidencia que la mortalidad en ambos géneros es decreciente, lo cual implica concluir sobre una esperanza de vida creciente para la población Colombiana.

Nota sobre los Datos

Los datos para realizar este análisis fueron tomados del texto *Colombia, Tablas Abreviadas de Mortalidad por Sexo Para Fechas Censales y Estimaciones Quinquenales 1995 - 2025* ([Gómez, 1998]) elaborado por el Dane bajo el marco de los estudios censales desarrollados luego del censo de 1993. Los registros de mortalidad no hacen ningún tipo de exclusión referente a enfermedades o causas de muerte particulares, por lo que se considera exhaustivo en todas las posibles causas registrables de mortalidad. Cabe mencionar que el estudio agrupa la información de los mayores de 80 años en el último grupo de edad (80+) el cual se presenta aquí como un único grupo, lo que puede afectar la varianza del modelo en este rango. En el análisis de residuales se observará cómo esta agrupación de datos afecta el modelo.

Se presenta, a modo de ilustración la información tomada del Dane ([Gómez, 1998]) para hombres y mujeres con corte de 1993.

4.1. Comparación entre modelos

4.1.1. Modelo LC ajustado a datos de hombres

Iniciando con la estimación del parámetro α_x observamos en el panel **I** de la figura **4-1** que las librerías `Demography` e `ilc` ajustan a valores muy similares, en estricto sentido solo se observa

x	t	$E_{x,t}(Hombres)$	$d_{x,t}(Hombres)$	$E_{x,t}(Mujeres)$	$d_{x,t}(Mujeres)$
1993	0	100,000.00	3,870.00	100,000.00	2,977.00
1993	1-4	96,130.00	1,279.00	97,023.00	1,098.00
1993	5-9	94,851.00	273.00	95,925.00	199.00
1993	10-14	94,578.00	326.00	95,726.00	189.00
1993	15-19	94,252.00	1,687.00	95,538.00	394.00
1993	20-24	92,565.00	2,812.00	95,144.00	465.00
1993	25-29	89,753.00	2,708.00	94,679.00	518.00
1993	30-34	87,045.00	2,439.00	94,161.00	613.00
1993	35-39	84,606.00	2,500.00	93,548.00	817.00
1993	40-44	82,106.00	2,407.00	92,731.00	1,111.00
1993	45-49	79,698.00	2,530.00	91,620.00	1,634.00
1993	50-54	77,168.00	3,334.00	89,986.00	2,521.00
1993	55-59	73,834.00	4,824.00	87,465.00	3,960.00
1993	60-64	69,010.00	6,323.00	83,505.00	5,618.00
1993	65-69	62,687.00	9,123.00	77,887.00	8,411.00
1993	70-74	53,564.00	11,119.00	69,476.00	11,532.00
1993	75-79	42,445.00	12,862.00	57,944.00	14,682.00
1993	80+	29,583.00	23,583.00	43,262.00	43,262.00

Tabla 4-1: Datos de expuestos al riesgo y muertes referentes al año 1993 para ambos generos ($E_{x,t}$ y $d_{x,t}$)

alguna diferencia relevante para los datos pertenecientes a las edades mayores. Por otro lado, la librería **gnm** tiende a desviarse hacia una menor fuerza de mortalidad media $\ln(\alpha_x)$ en las edades entre 0 – 40 y a tener una mayor magnitud en las edades de 40 – 80 años, lo que en primera instancia nos indica que la librería **gnm** logra representar una mayor variabilidad en estos rangos extremos de edades. Para el caso de la estimación de la sensibilidad a los cambios en t , β_x , observamos en el panel **II** de la figura **4-1** una diferencia significativa entre los ajustes presentados por las dos primeras librerías en el sentido que para la edad $x = 80$ la librería **Demography** continua una tendencia decreciente, sin embargo la librería **ilc** cambia de tendencia presentando un incremento en la sensibilidad, lo que se podría anotar como una deficiencia en la librería **Demography** ya que no logró captar este cambio de dirección en el parámetro. En el caso de la librería **gnm** se nota un cambio menor para β_x en los años menores de 20 y una dinámica diferente a las demás librerías para los mayores de 55 años. Es importante anotar que las tres librerías coinciden en poca variabilidad de este parámetro entre las edades de 20 a 60 años, los que sugiere que el valor de la tasa de mortalidad para estas edades es invariante ante cambios en el tiempo medidos a través de κ_t (recordemos la ecuación 2-6, $m_{x,t} = e^{\alpha_x + \beta_x \kappa_t + \epsilon_{x,t}}$). Por su parte, κ_t presenta una pendiente mayor para la librería **gnm**, nótese el cambio de pendiente entre los años 1993 y 1999, a diferencia de las

librerías **Demography** e **ilc** las cuales se muestran con una pendiente constante. Dado que los parámetros estimados no son observables, no hay forma de comparar con la realidad, sin embargo, realizando una revisión bibliográfica, podemos corroborar que los parámetros estimados a través de **gnm** son los que más se asimilan en tendencia a los datos presentados en [Gómez, 1998].

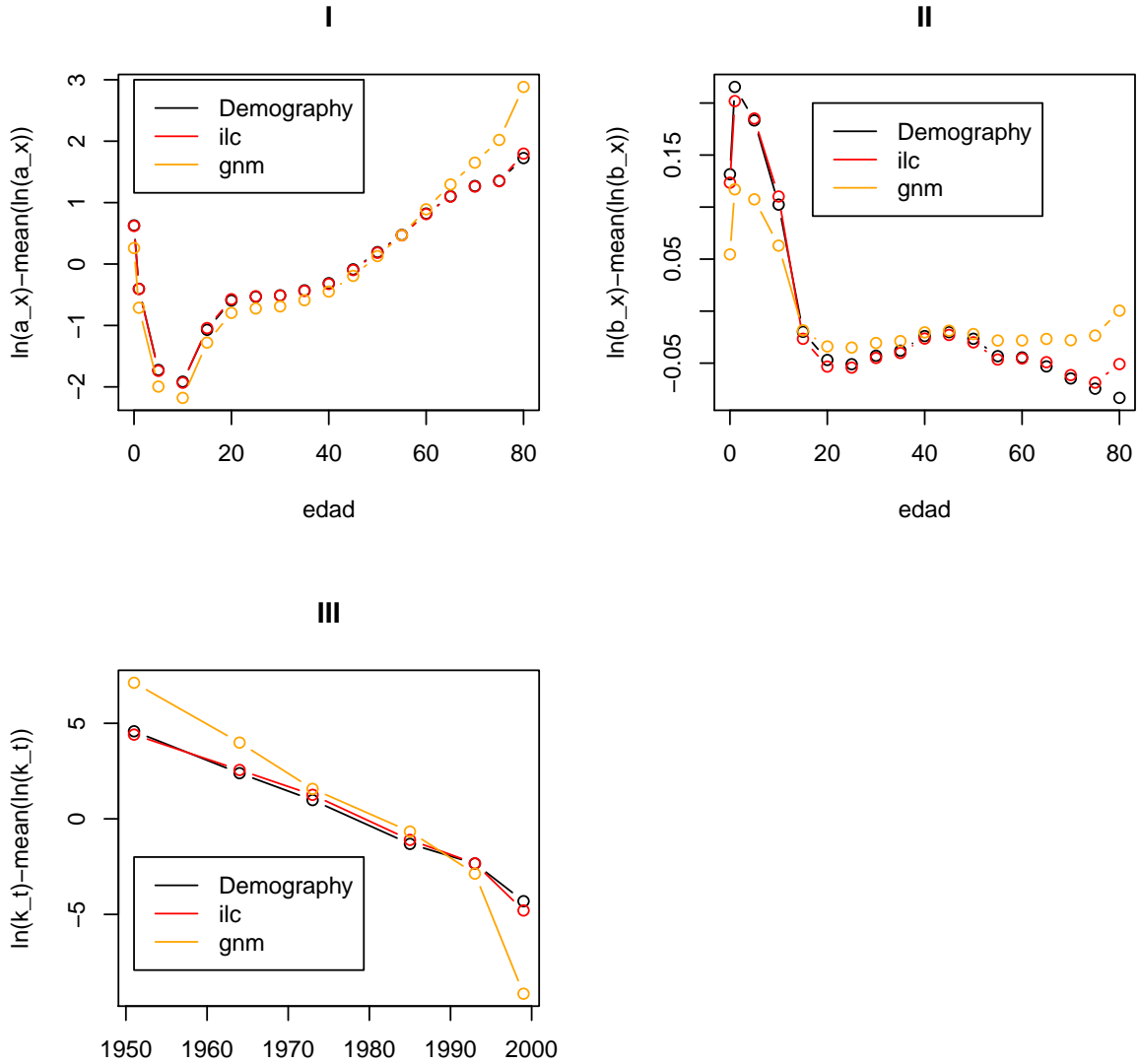


Figura 4-1: Comparación entre parámetros α_x , β_x y κ_t para los modelos Demography, ilc y gnm para hombres. Para cada uno de los parámetros α_x , β_x y κ_t se presenta normalizado restando su respectiva media en aras de poder hacer comparaciones ya que los valores generados con la librería gnm presentan diferencias debido a su algoritmo de estimación

En la figura 4-2, 4-3 y 4-4 se observa la evolución de $\mu_{x,t}$ a través de los años estimados para cada una de las librerías analizadas. Claramente se observa la tendencia decreciente en la mortalidad a medida que pasa el tiempo. El punto común entre las librerías se encuentra en las edades entre 0 y 20 años, para lo cual se presenta una caída con curvatura y evolución (en magnitud) similar. A partir del punto de inflexión común que se presenta en los 20 años, se presentan diferencias entre las librerías usadas para la estimación. En la figura 4-2 y 4-3 observamos que se presentan pocas diferencias entre las librerías **Demography** e **ilc**. Luego de una revisión numérica, las mayores diferencias se presentan en la estimación de $\mu_{x,t}$ para $x \geq 70$. La librería **Demography** presenta una evolución creciente del parámetro para esta edad, lo cual se opone a la sugerencias observadas en la librería **gnm** (figura 4-4) en la que la mortalidad es decreciente a medida que el tiempo avanza. En contraste, la librería **ilc** no captura el cambio en la mortalidad en este punto.

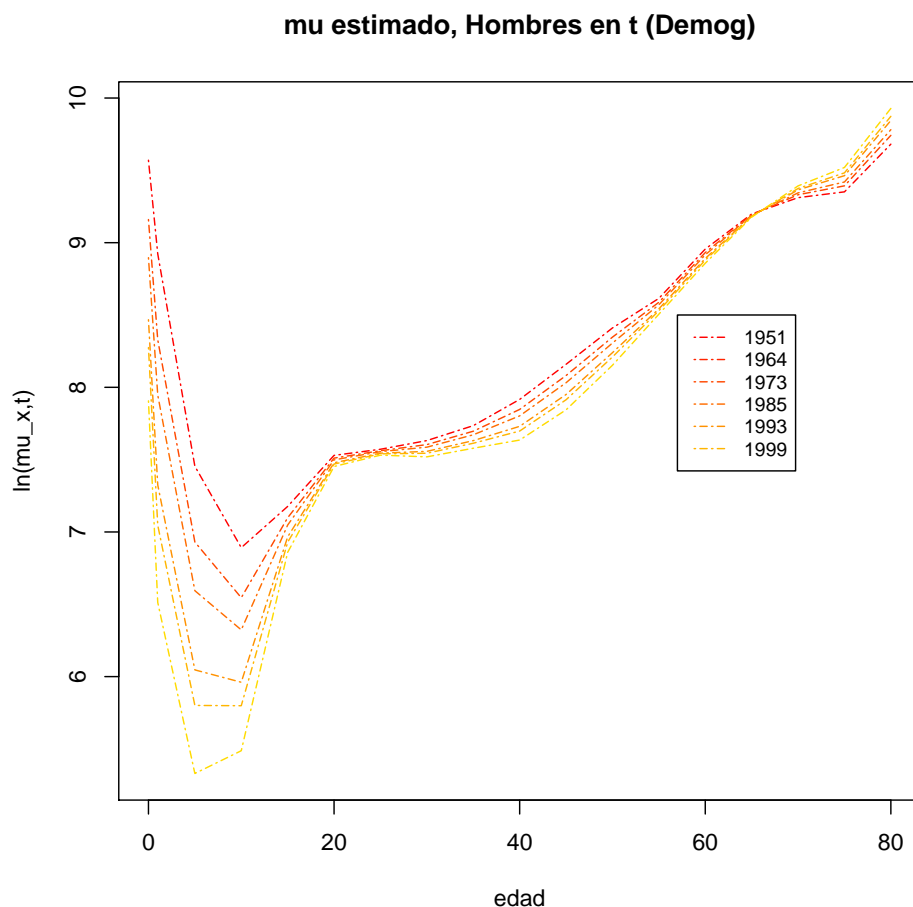


Figura 4-2: $\mu_{x,t}$ estimado para hombres con librería Demography

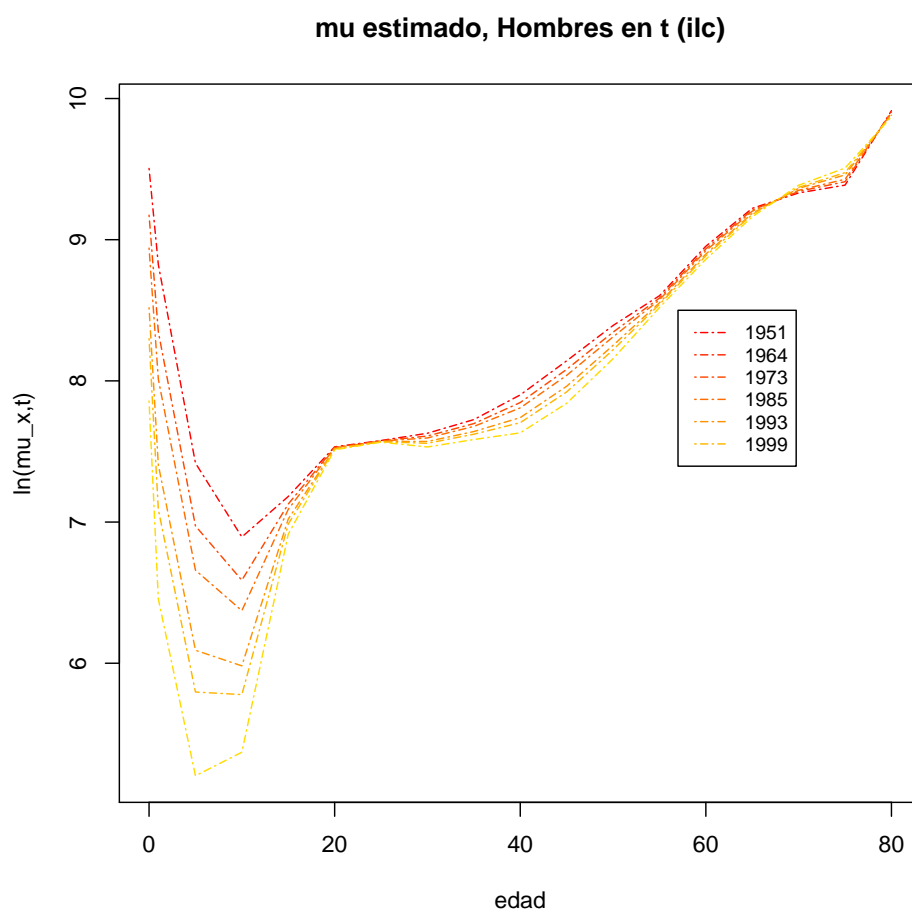


Figura 4-3: $\mu_{x,t}$ estimado para hombres con librería ilc

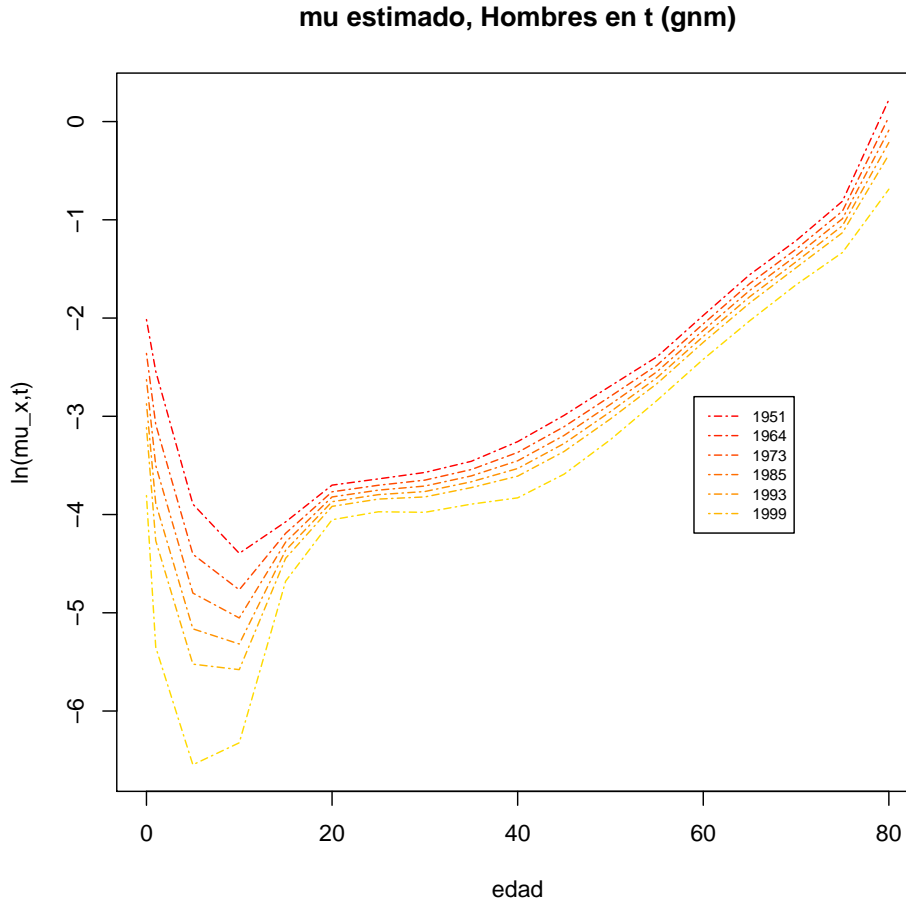


Figura 4-4: $\mu_{x,t}$ estimado para hombres con librería gnm

4.1.2. Modelo LC ajustado a datos de mujeres

En la figura 4-5 se ilustra la estimación de los diferentes parámetros del modelo LC para las tres librerías analizadas. De nuevo para el caso de α_x observamos similitud en la estimación, con solo una notoria diferencia para las edades superiores a 60 años en el caso de la librería **gnm**, la cual presenta una cifra mas alta para las últimas edades. En el panel **II** vemos la evolución de β_x estimado, en donde se nota un cambio más significativo entre las sensibilidades presentadas por las librerías **Demography** e **ilc**. La estimación de este parámetro bajo **gnm** presenta una menor variabilidad a lo largo de todas las edades, siendo particularmente constante para los 50 años en adelante. De forma similar a los hombres, se anota la similitud que presentan las tres librerías para el rango de edad entre los 20 y 60 años, lo que repercute directamente en la tasa de mortalidad y su respectiva variación a través del tiempo (ver la ecuación 2-6). La estimación de la evolución de κ_t en las mujeres evidencia una caída similar a la de los hombres. Cabe anotar la reversión en la tendencia presentada por la librería **ilc**

para el año 1999, la cual podría afectar significativamente los pronósticos que se realicen de la mortalidad.

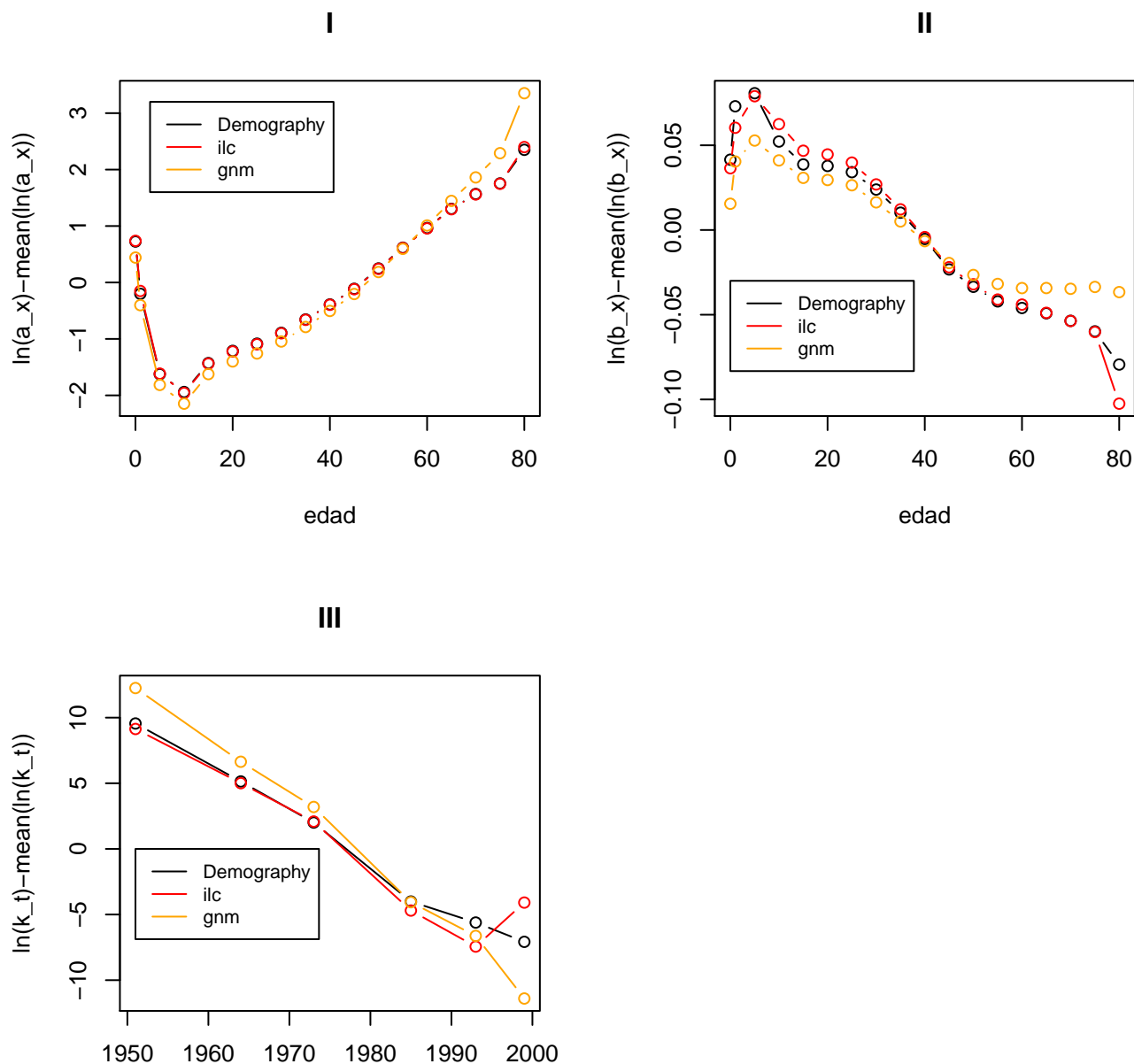


Figura 4-5: Comparación entre parámetros α_x , β_x y κ_t para los modelos Demography, ilc y gnm para mujeres. Para cada uno de los parámetros α_x , β_x y κ_t se presenta normalizado restando su respectiva media en aras de poder hacer comparaciones ya que los valores generados con la librería gnm presentan diferencias debido a su algoritmo de estimación

Para el caso de las mujeres el parámetro $\mu_{x,t}$ tiene un comportamiento similar para las primeras dos librerías (figuras 4-6 y 4-7), siendo tal como el caso de los hombres, en el que se diferencian por la estimación para la última edad. En esta estimación las librerías **Demography** e **ilc** coinciden en asignar una mortalidad creciente para la edad de 70 años luego de haberse observado mortalidad decreciente entre 0 y 70 años. También se nota una diferencia para la edad de 20 años, en la que la librería **ilc** tiende a marcar más el cambio en la mortalidad entre los años 1993 y 1999. En contraste, el modelo estimado con la librería **gnm** (figura 4-8) evidencia una evolución de la mortalidad a través del tiempo más ajustada a las revisiones hechas por autores como [Currie, 2014] y [Betzuen, 2010] en las que se observan patrones más suaves de cambio como particularmente se observa para edades como $x = 60 - 65$.

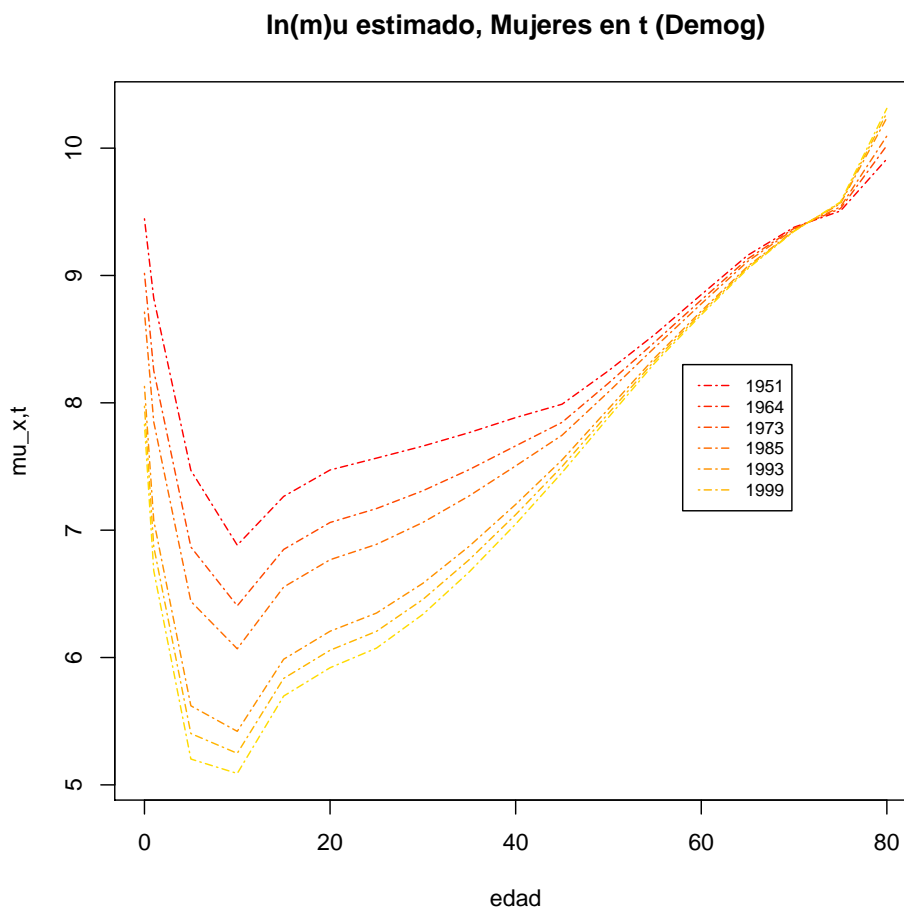


Figura 4-6: $\mu_{x,t}$ estimado para mujeres con librería **Demography**

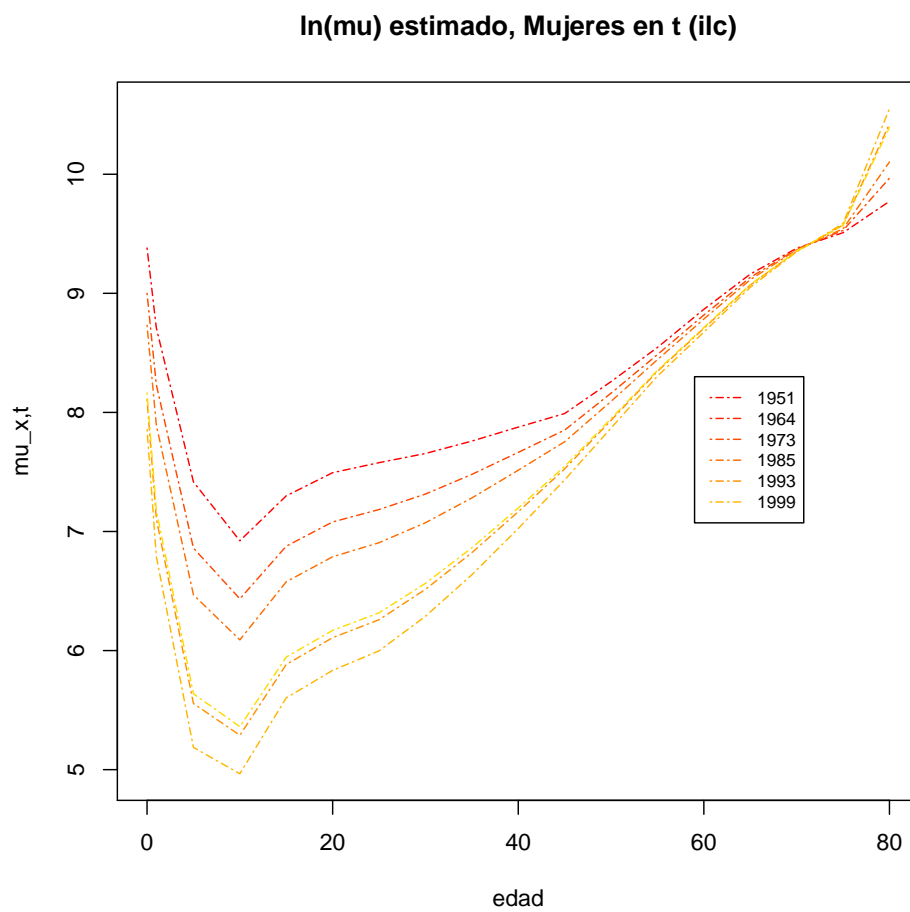


Figura 4-7: $\mu_{x,t}$ estimado para mujeres con librería ilc

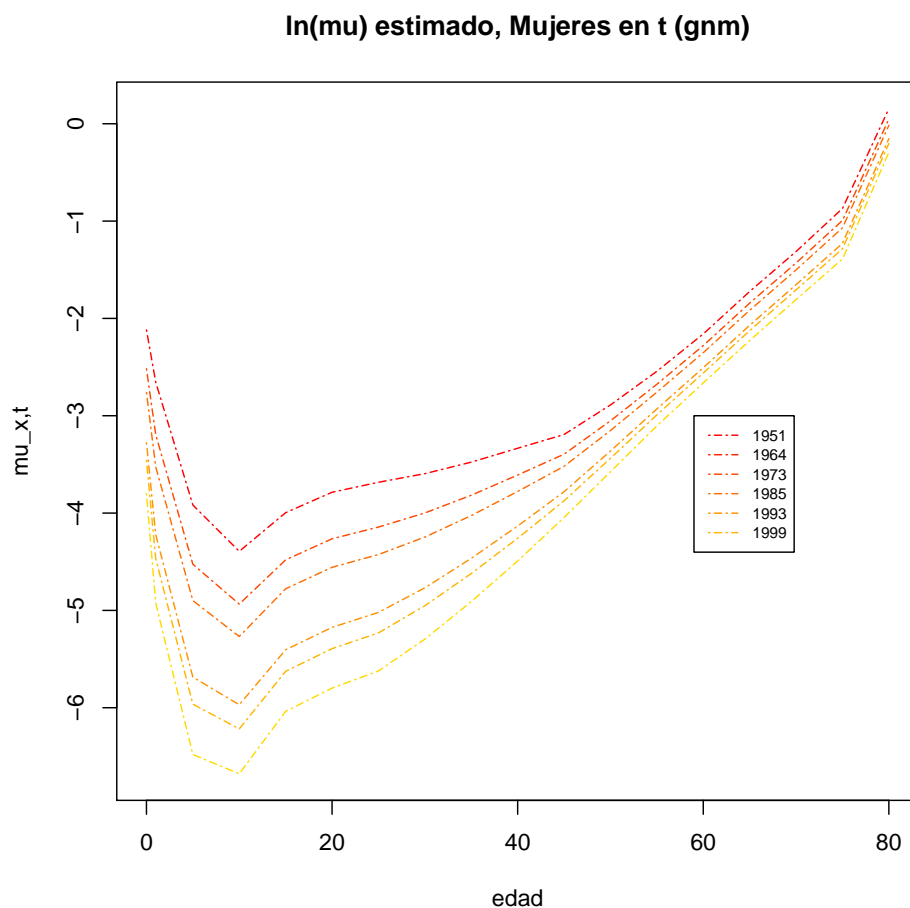


Figura 4-8: $\mu_{x,t}$ estimado para mujeres con librería gnm

En la tabla 4-2 y 4-3 se presentan los residuales tipo Deviance y tipo Pearson de la estimación de $\hat{d}_{x,t}$ (Varianza del error residual para $\hat{d}_{x,t} = E_{x,t}e^{\hat{\alpha} + \hat{\beta}\kappa}$) bajo las tres metodologías planteadas, es decir, librería **Demography** que usa descomposición singular (SVD), librería **ilc** a través de máxima verosimilitud (GLM) y la librería **gnm** que minimiza el estadístico Deviance (GNM).

Género	SVD	GLM	GNM
Hombres	13,881.7	14,748.9	8,540.7
Mujeres	23,454.2	17,999.7	10,733.0

Tabla 4-2: Residuales tipo Deviance de la estimación de $\hat{d}_{x,t}$ discriminados por género

Género	SVD	GLM	GNM
Hombres	14,145.087	14,377.569	8,619.263
Mujeres	22,996.867	15,258.976	10,199.326

Tabla 4-3: Residuales tipo Pearson de la estimación de $\hat{d}_{x,t}$ discriminados por género

Con base en los residuales de la tabla 4-2 y 4-3 en los que se observa menor variación del error tipo Deviance y Pearson entre $d_{x,t}$ y $\hat{d}_{x,t}$ para ambos géneros, concluimos que la mejor estimación es la realizada por el modelo log-bilineal de Poisson a través del uso de la herramienta `gnm`.

Como complemento al análisis se presenta la figura 4-9 en la que se exponen los residuales tipo Deviance para hombres y mujeres. Para el caso de los hombres por edad (panel I) se evidencia heterocedasticidad, principalmente en las edades entre los 0 y 15 años y para el ultimo corte de edad de 80 años. En las edades menores esto se justifica por la dinámica del rango en el que la mortalidad cambia mucho más con referencia al resto de edades. En la edad de 80 años, la diferencia en la estimación vs observado responde a baja calidad en los datos ya que esta edad representa el rango de edad de 80 años en adelante, la cual representa un número mayor de muertes que sus intervalos previos.

En el panel II se encuentra la figura de residuales para hombres por año. En la gráfica se observan algunos outliers, sin embargo no afectan la homocedasticidad presentada a través de los años.

Para el género femenino se presenta homocedasticidad a lo largo de las edades, exceptuando el caso del último rango de edad (80), el cual, de forma similar al de los hombres, expresa un rango de datos para 80 o más años. Finalizando el análisis, en el panel IV de la gráfica 4-9 se observan los residuales por año en los que, a pesar de algunos outliers, la varianza tiene un comportamiento deseable para realizar pronósticos.

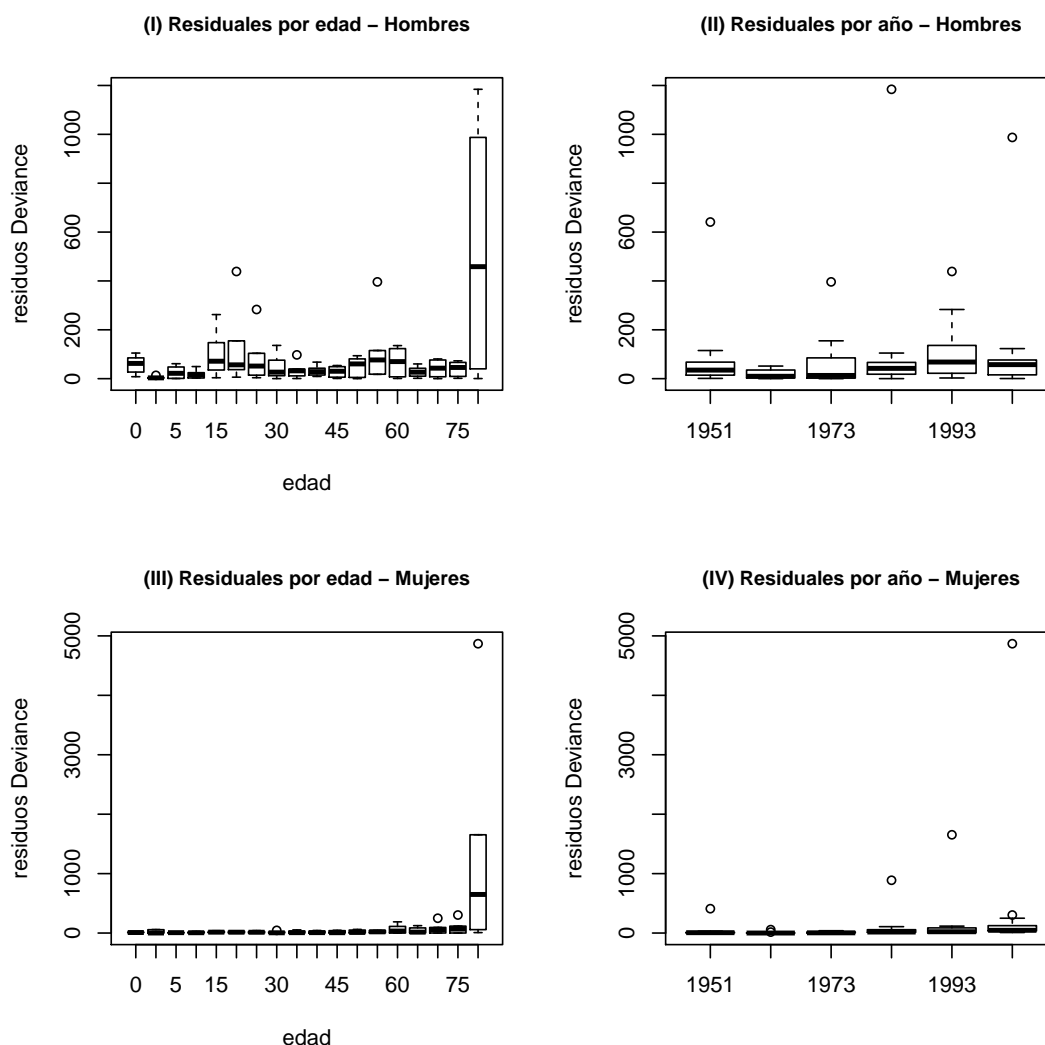


Figura 4-9: Residuales Deviance para el modelo gnm por edad y año en hombres y mujeres

4.2. Estimación del modelo a través de gnm

Luego de haber definido la mejor metodología de estimación del modelo LC para los datos poblacionales de Colombia, procedemos a realizar una proyección de la mortalidad en aras de ilustrar la tendencia que este indicador muestra a través del tiempo.

Se presenta la forma como se realiza la estimación y proyección del modelo LC en R a través de la librería **gnm** siguiendo las recomendaciones de [Currie, 2014]. El ajuste a los parámetros se sugiere para aplicar las restricciones dadas en 2-7.

```
## Modelo LC

LCgnm <- gnm(mDeaths ~ -1+ Age + Mult(Age, Year),
  offset = log(mExposure), family = poisson(link="log"), data = dane)

coeff=LCgnm$coefficients

# Ajuste parametros Currie 2014

alfa=coeff[1:nx]
beta=coeff[(nx+1):(2*nx)]
k=coeff[(2*nx+1):(2*nx+nt)]

kR = sum(k)/nt
betaR = sum(beta)/nx

alfa.est = alfa +kR*beta.h
k.est = nx*betaR*(k-kR)
beta.est = beta/(nx*betaR)
```

En las tablas 4-4 y 4-5 se presentan los valores estimados de los parámetros α_x , β_x y $\hat{\kappa}_t$.

Edad	$\hat{\alpha}_x$	$\hat{\beta}_x$
0	-2.99	0.07
1	-3.83	0.10
5	-5.25	0.11
10	-5.58	0.10
15	-5.05	0.09
20	-4.83	0.09
25	-4.69	0.08
30	-4.48	0.07
35	-4.22	0.06
40	-3.93	0.05
45	-3.63	0.04
50	-3.24	0.03
55	-2.83	0.02
60	-2.42	0.02
65	-1.98	0.02
70	-1.57	0.02
75	-1.14	0.02
80+	-0.08	0.02

Tabla 4-4: Valores de $\hat{\alpha}_x$ y $\hat{\beta}_x$ a través de la librería gnm

t	1951	1964	1973	1985	1993	1999
$\hat{\kappa}_t$	12.25	6.64	3.19	-4.06	-6.63	-11.40

Tabla 4-5: Valores de $\hat{\kappa}_t$ a través de la librería gnm

4.3. Proyección de mortalidad

El objetivo del modelo LC es realizar proyecciones de la tasa central de mortalidad a partir de la modelación del parámetro κ_t como una serie de tiempo. En este caso, la serie de tiempo se estimó con un modelo semi-paramétrico de regresión local Loess con un ancho de ventana de 0.75 (sugerido por defecto). La metodología no paramétrica Loess es ideal debido a que se requiere de un modelo que le asigne una importancia mayor a las observaciones cerca de la fecha de estimación. En dicho modelo cada punto se determina por una línea de regresión en donde se realiza una ponderación de las observaciones en función de la distancia desde el respectivo punto. Ver un ejemplo de aplicación de este tipo de modelo en regresión de tasas de mortalidad en: [Miller, 2011].

La proyección en **R** se realizó de la siguiente manera:

```
#----análisis pronósticos con Loess span=0.75, Hombres

kt.h=ts(k.h.est,frequency=1)

m.lo.gnm.h=loess(kt.h ~ t,
control = loess.control(surface = "direct"))

tt = seq(1999,by=4,length.out=6)

pr.lo.gnm.h = predict(m.lo.gnm.h,
data.frame(t = tt))
```

En la figura **4-10** podemos observar los valores estimados para los años 1951, 1964, 1973, 1985, 1993 y 1999 y proyectados para los años 2003, 2007 y 2011 para hombres y mujeres. En esta figura se evidencia una diferencia importante entre la tendencia de la mortalidad entre los géneros.

El objetivo planteado en este texto, de modelar y proyectar la dinámica de la mortalidad poblacional se evidencia en este aparte del texto. La gráfica sugiere una disminución en el índice de mortalidad para ambos géneros, siendo más marcada esta disminución para el género masculino, es decir, los hombres presentan una mortalidad menor que las mujeres para los periodos mayores al año 2000. Esta evidencia nos sugiere un argumento que se busca evidenciar con este texto; la esperanza de vida en la población analizada está incrementando con el tiempo, o de otra manera, la dinámica demográfica de Colombia está cambiando hacia una mayor longevidad en ambos géneros.

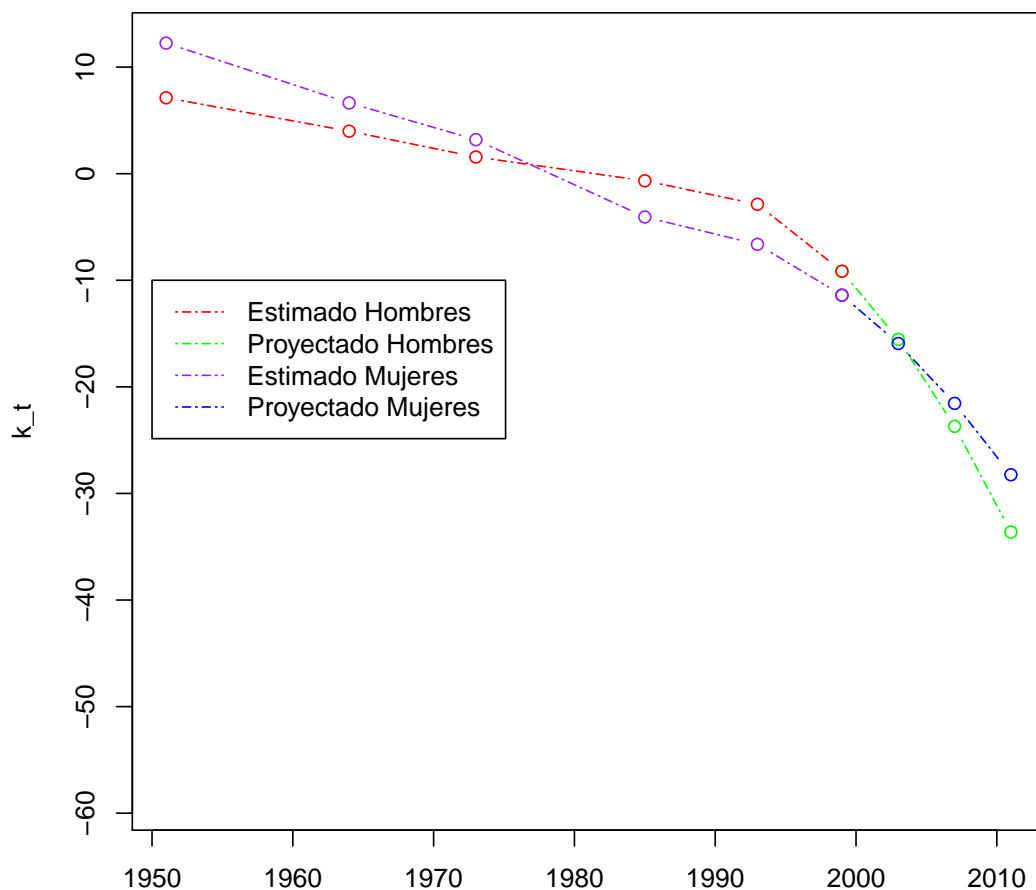


Figura 4-10: k_t estimado (1951-1999) y proyectado (2003-2011)

Una vez estimado y proyectado el parámetro κ_t se procede a estimar la tasa central de mortalidad para cada edad a través del tiempo. En la figura 4-11 se observa la evolución de esta tasa estimada para los hombres a lo largo del tiempo. Lo primero que llama la atención es la forma marcada en la que evoluciona la mortalidad para las edades entre los 0 y 20 años, evidenciando el carácter exponencial del modelo llevando la mortalidad a unos niveles mínimos alejados de lo que sucede en los periodos siguientes. Lo anterior es seguido de una evolución mas constante a partir de los 40 años. Además, el modelo logra capturar un cambio importante para una edad de aproximadamente 70 años en donde se presenta un punto de inflexión para la estimación del 2011. Se sabe a partir del análisis de residuales que, los datos para las edades de 80 años y más presentan sobredispersión, la cual puede ser la causante de la mayor variación en la mortalidad para este rango, como fue anotado anteriormente esto

es causado por la agrupación que tienen los datos para el rango de 80 años y más, ver tabla 4-1.

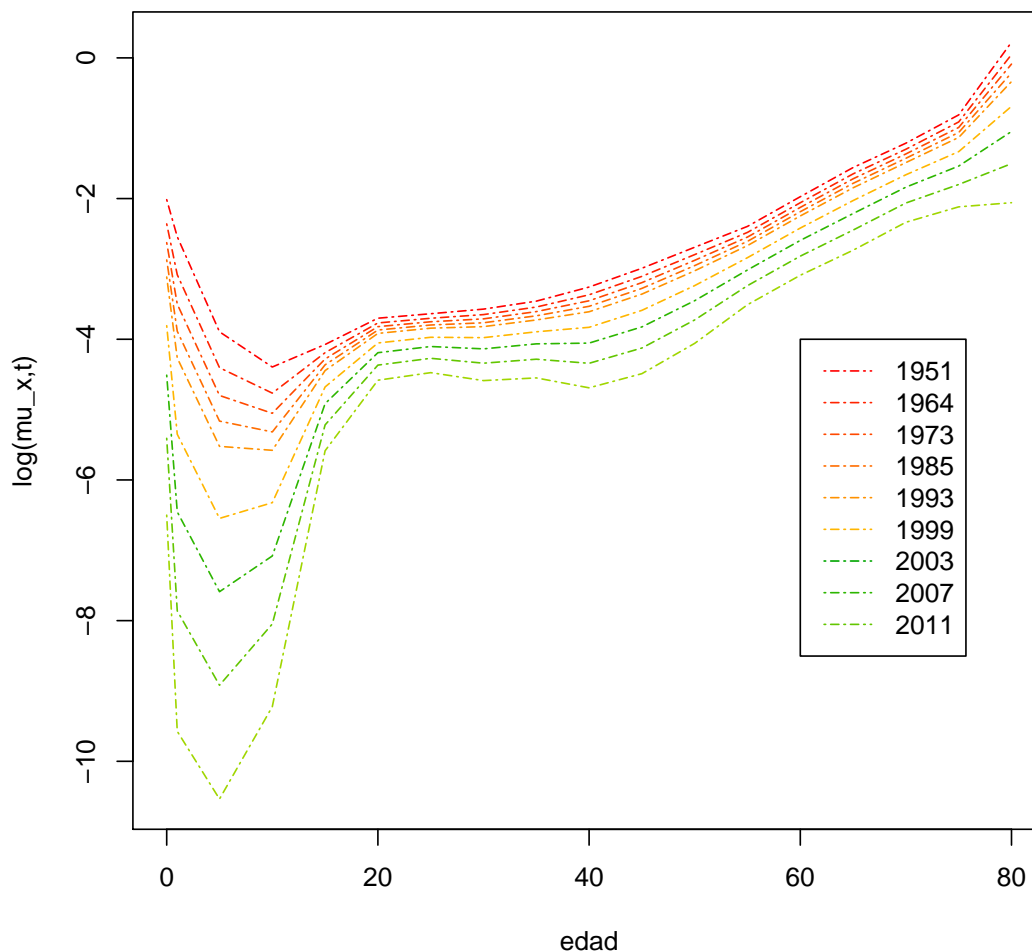


Figura 4-11: $\log(\mu_{x,t})$ estimado (1951-1999) y proyectado (2003-2011) - Hombres

Para el caso de las mujeres observamos la figura 4-12. En esta se evidencia una variación más distribuida a lo largo de las edades entre los 0 y 50 años, seguida por una menor variación para las edades mayores a 50.

En contraste podemos afirmar que existen diferencias importantes a lo largo de las edades entre los hombres y mujeres, principalmente en el rango de los 20 a los 50 años en el que las mujeres de esta edad presentan una variación más importante en el logaritmo de la tasa central de mortalidad. De forma similar a los hombres, se evidencia un punto de quiebre

para una edad de aproximadamente 70 años, que en este caso define una dinámica menos decreciente de la tasa analizada.

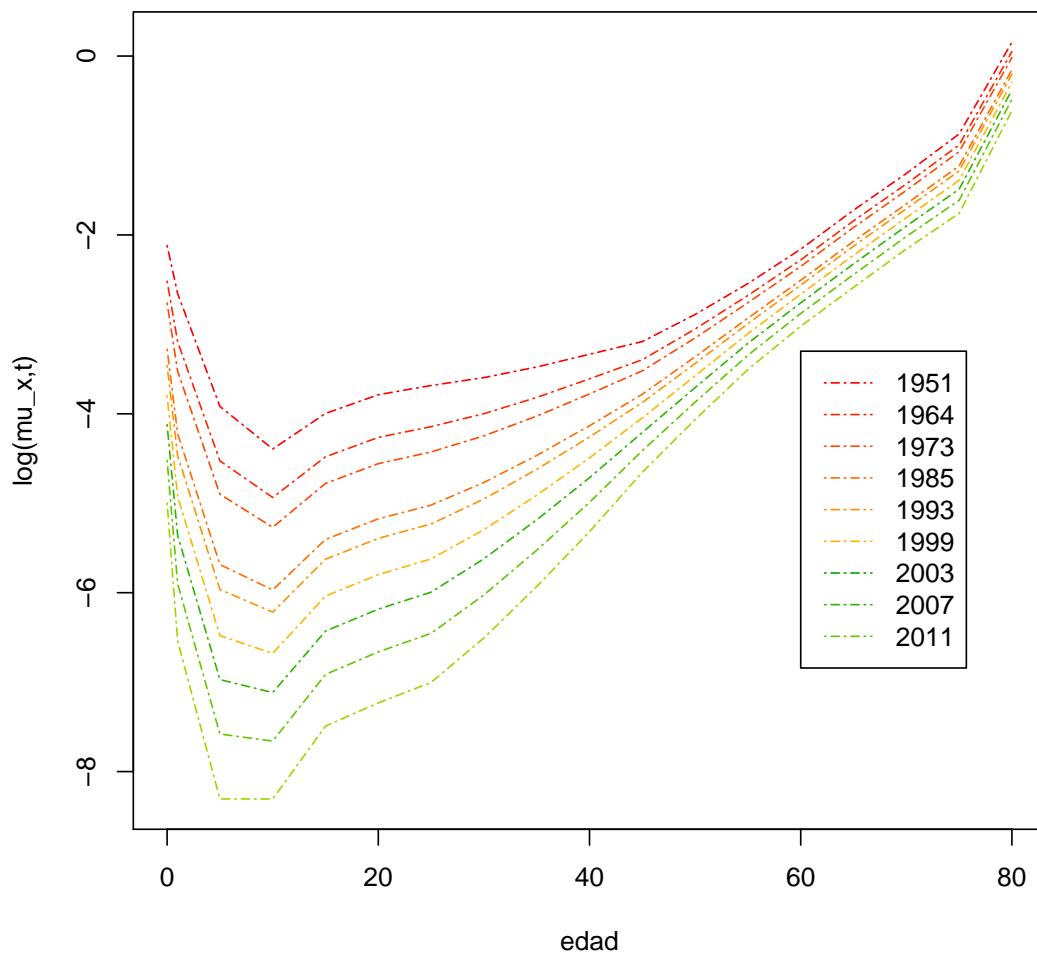


Figura 4-12: $\log(\mu_{x,t})$ estimado (1951-1999) y proyectado (2003-2011) - Mujeres

5 Estimación de la población universitaria

En este capítulo se ilustra una aplicación adicional del modelo LC en el que se modela la población estudiantil de pregrado en la Universidad Nacional de Colombia sede Medellín a través de 34 cohortes. La modelación de estos datos se motiva en la necesidad de implementar políticas académicas y reglamentarias en la universidad en aras de reducir la mortalidad académica entre los estudiantes. Siguiendo la línea del capítulo pasado, se ajusta el modelo LC a partir de la librería `gnm` de [Turner & Firth, 2015] para obtener los valores de κ_τ . Luego se realiza una proyección mediante dos metodologías y se comparan. Una forma es la proyección a través de la metodología Box-Jenkins con un ARIMA (1, 1, 0) y la otra metodología usada es la misma que se implementó para proyectar mortalidad poblacional: la regresión no-paramétrica Loess. En ambas estimaciones se observó un patrón aleatorio del parámetro κ_τ durante las cohortes comprendidos entre el 1989 – 02 y 2006 – 01, para 34 semestres (o cohortes). Con la estimación de κ_τ se realizaron 10 pronósticos para las cohortes entre 2006 – 02 y 2011 – 01 bajo las dos metodologías y se compara el (último) pronóstico con la información conocida hasta el octavo semestre de esta cohorte, correspondiente a la información conocida en el 2014. Es válido anotar que la Universidad Nacional tuvo una reforma al sistema académico la cual entró a regir en el semestre 2008 – 01. Dicha reforma incorporó el sistema de créditos y cambió las condiciones de permanencia de los estudiantes en la universidad.

Teniendo en mente que el periodo a proyectar es post reforma, se podrá ver los efectos iniciales en la población universitaria.

Se inicia el análisis con la figura 5-1 en la que se resume la evolución de cada cohorte, es decir, cuantos semestres cursó el estudiante de dicha cohorte antes de retirarse de la universidad. En esta gráfica se observa una disminución importante en la población estudiantil en los primeros semestres, también se observa un punto de quiebre natural entre el semestre 10 y 11 y finalmente se nota una masa importante de población universitaria para semestres superiores al décimo.

Nota sobre los Datos

En este trabajo se consideraron datos de pregrado para las cohortes entre 02/1989 y 01/2006, para la Universidad Nacional de Colombia, Sede Medellín, proporcionados por el Sr. Fabio

Rico de la Oficina de Planeación Académica de la Sede. Los datos presentan algunas inconsistencias, que se analizan a continuación. Por su definición la función l_x debe ser monótona decreciente, $l_{x+1} \leq l_x$; pero esto no sucede siempre, posiblemente debido a reingresos y traslados. Por tanto, se aplicó un suavizamiento monótono a los datos, mediante la función *Loess* de R. Adicionalmente, los datos de cada semestre en los últimos periodos tienen valores pequeños ó tienen valores cero. Esto genera valores L_x muy pequeños que producen estimadores d_x/L_x con mucha varianza. La acción que se tomó fué considerar en cada semestre solamente los valores x tales que $x \leq 17$, es decir, se utilizaron datos de supervivientes hasta 17 semestres; quienes presentaron una permanencia de más de 17 semestres se distribuyen de manera irregular, con poco datos.

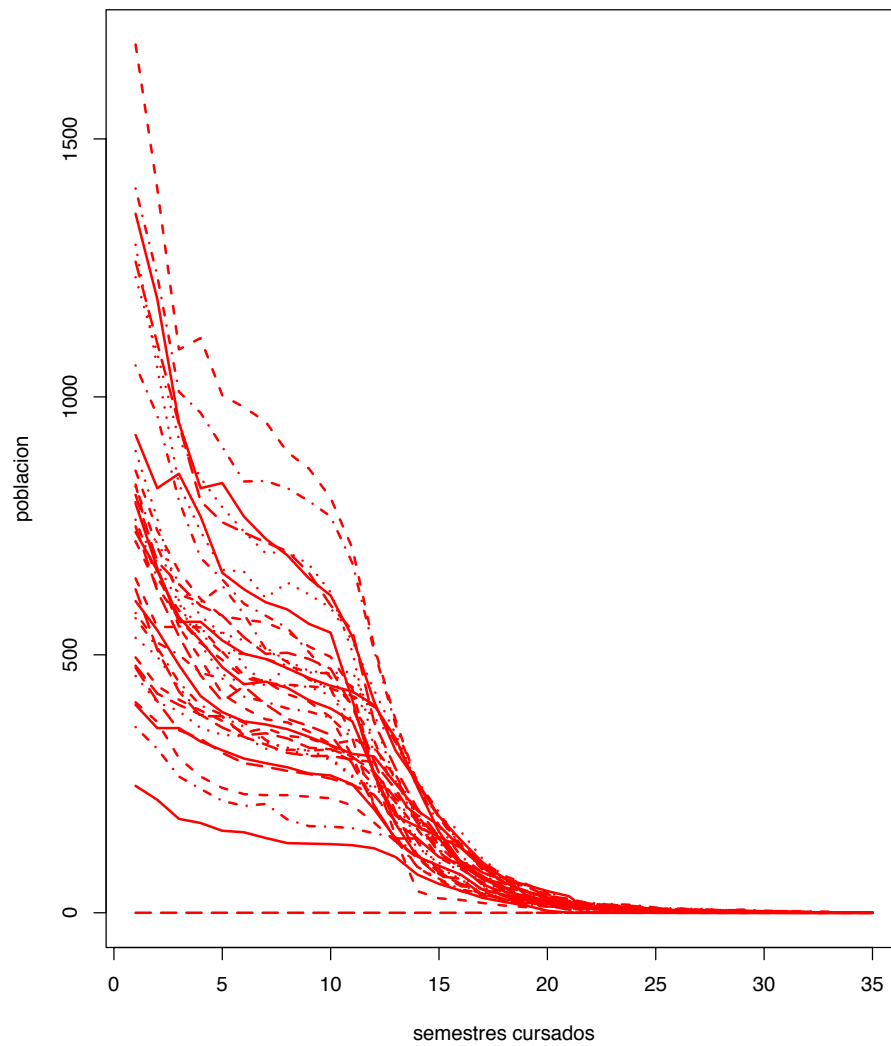


Figura 5-1: Evolución de los semestres: 1989/02 - 2006/01

5.1. Estimación del modelo a través de gnm

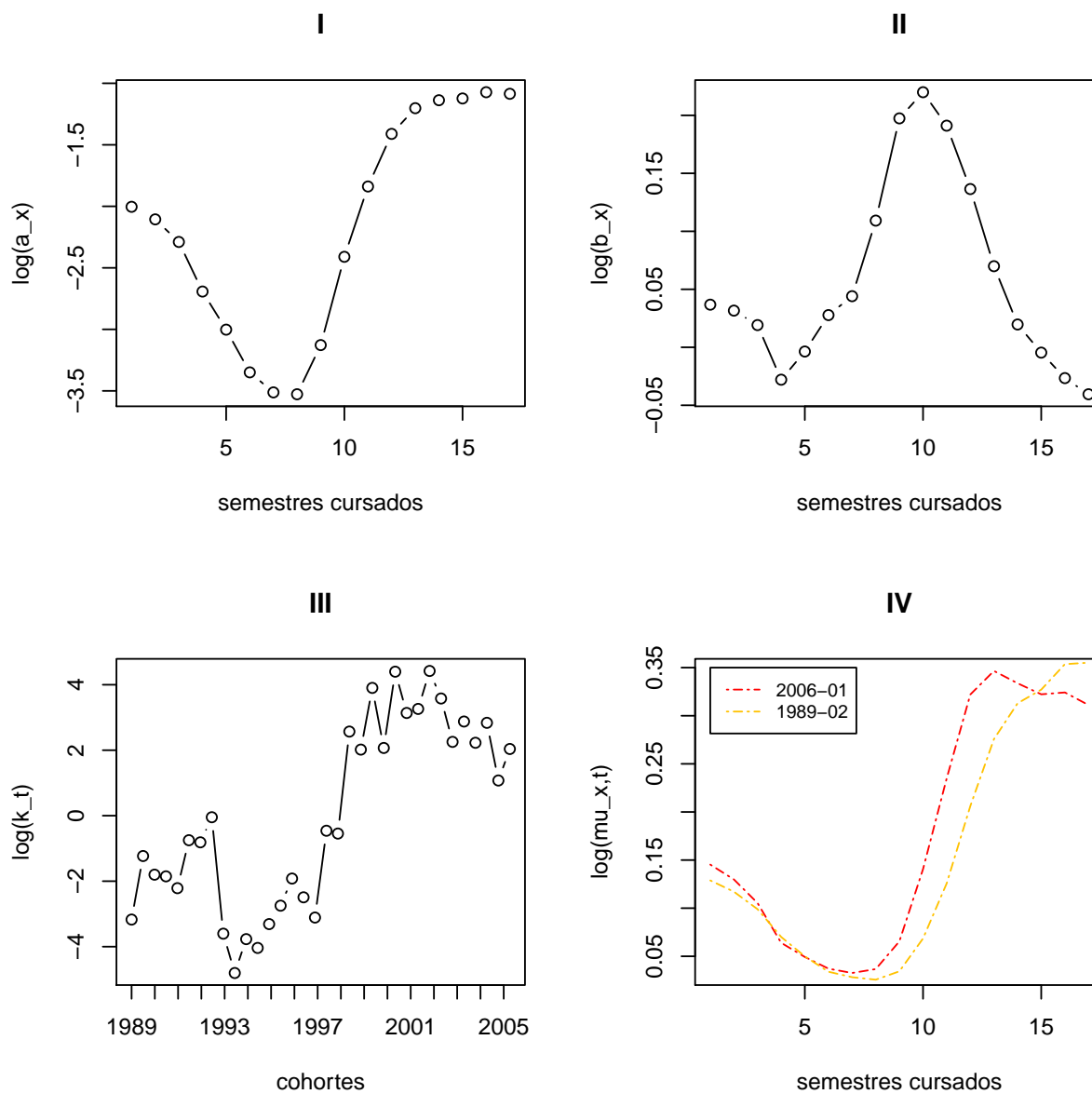


Figura 5-2: Modelo LC para mortalidad académica, parámetros α_x , β_x , κ_τ y $\log(\mu_{x,\tau})$ estimados por Max.Ver.

Siguiendo la línea presentada en la sección 4.2 se realiza la estimación de los parámetros α_x , β_x , κ_τ , y $\mu_{x,\tau}$ ($x = 1, 2, 3, \dots, 17$ y $t = 1989-02, 1999-01, 1999-02, \dots, 2005-02, 2006-01$). Los resultados pueden ser observados en la figura **5-2**. La salida de alumnos se evidencia en el panel **I** con el parámetro α_x , para lo cual se observa una menor deserción en los semestres 6, 7 y 8 para luego tomar una inclinación positiva y llegar a un equilibrio luego del semestre número 13, en donde se presenta la mayor relación con la fuerza de mortalidad. En el panel **III** se observa la evolución de κ_t cuya dinámica muestra una reversión a mediados del 2005 al nivel de inicios de la década de 1990. $\mu_{x,\tau}$, quien determina la tasa central de mortalidad es mostrada en el panel **IV** para los datos correspondientes a 1989 – 02 y 2006 – 01. Se observa en la figura que la tasa es mayor para la cohorte 2006 – 01, lo que permite concluir que la deserción ha aumentado para los estudiantes que permanecen hasta tres semestres en la universidad. Para los datos correspondientes a los semestres 4 y 5 la tasa ha permanecido constante para luego cambiar notoriamente entre los estudiantes que permanecen entre 7 y 15 semestres, para los cuales la deserción estudiantil se ha incrementado. Para los semestres siguientes la tendencia se revierte.

En la figura **5-3** se presenta la evolución de la dinámica poblacional para cada uno de los 17 semestres analizados a través de las cohortes. En otras palabras, se analiza para un semestre en específico cómo fue su evolución a través del tiempo. Revisando un caso particular, en el semestre 1, el cual representa los estudiantes que ingresaron a la universidad y permanecieron en ella solo un semestre, evoluciona de manera creciente a través de las diferentes cohortes, esto es, en una cohorte del 2004, un estudiante tiene más probabilidades de salir en el primer semestre que para una cohorte del 1990. También se puede evidenciar, que los estudiantes que permanecen entre 3 y 9 semestres son los más constantes entre las cohortes. Como punto relevante se evidencia que los alumnos que están 10 o mas semestres presentan un cambio notorio entre las cohortes de 1993 y 1998, con un incremento en variabilidad hasta las últimas cohortes revisadas. Respecto a la reforma mencionada, la cual inició en el semestre 2008 – 01, se considera que puede afectar a alumnos que ingresaron alrededor del año 2001, quienes para el año 2008 llevaban aproximadamente 14 semestres en la universidad. La figura **5-3** evidencia a partir del año 2001 una disminución en variabilidad para los estudiantes que permanecen entre 8 y 16 semestres en la universidad y una tendencia de disminución en el logaritmo del parámetro $\mu_{x,\tau}$, lo cual podría ser un efecto de la reforma.

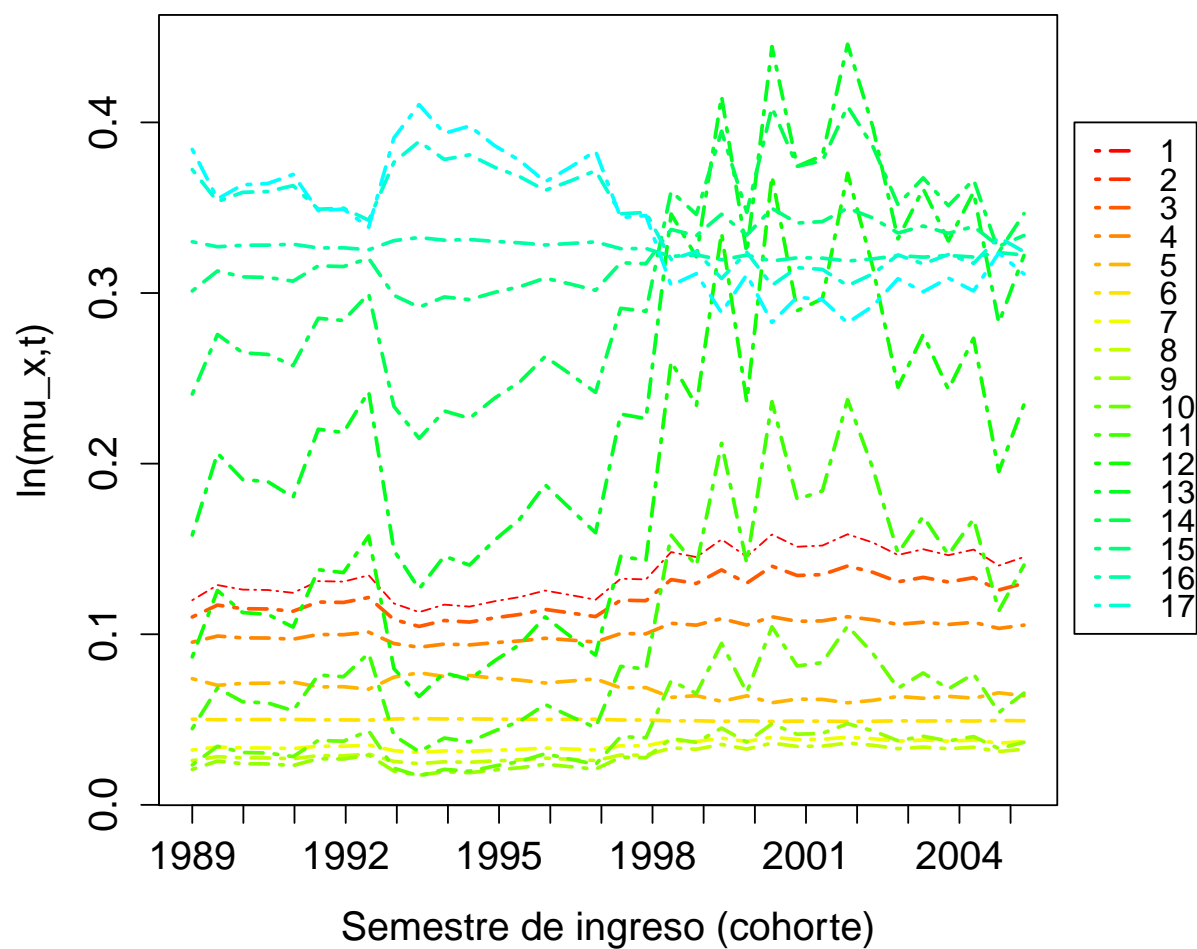


Figura 5-3: $\log(\mu_{x,t})$ para los 17 semestres a través de las 34 cohortes analizadas

En la figura 5-4 se encuentran los residuos de la estimación:

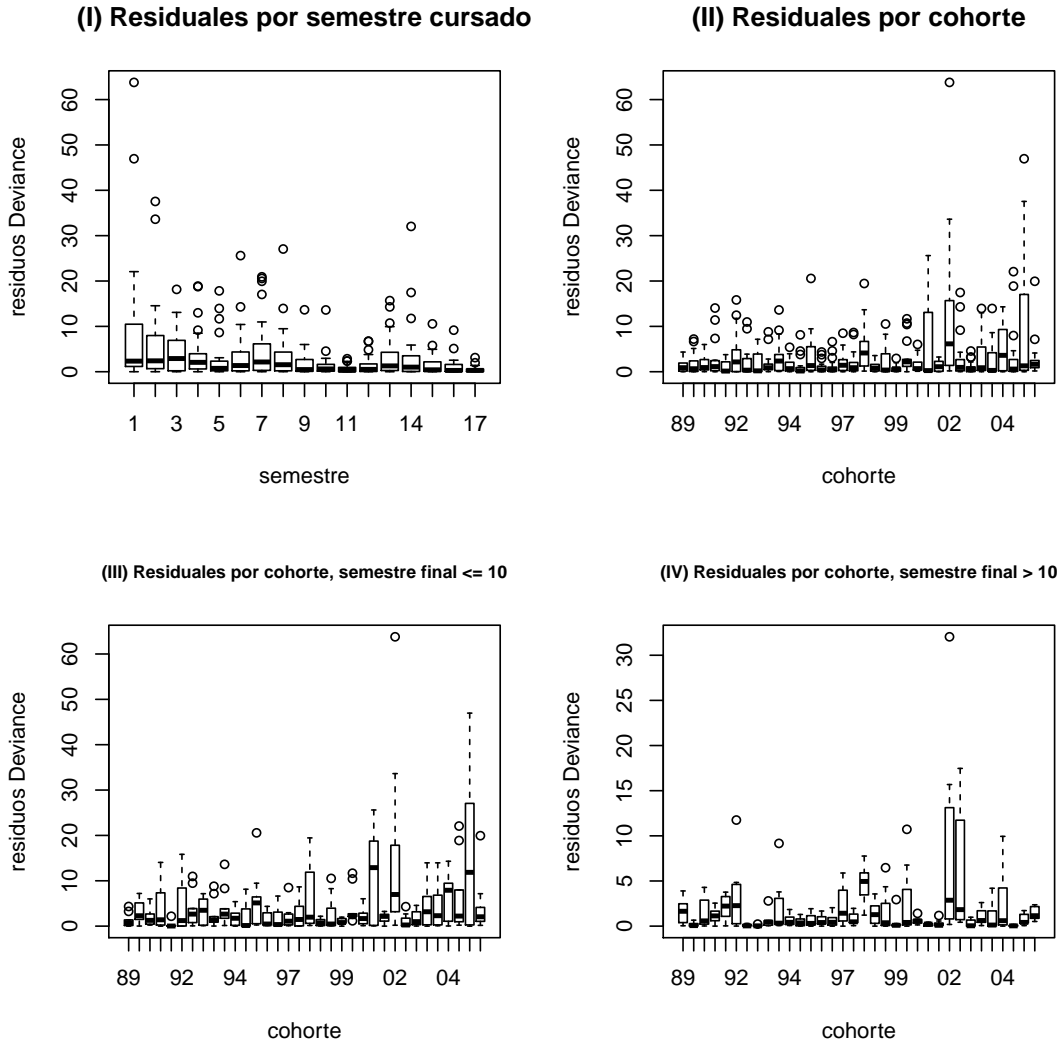


Figura 5-4: Residuos (deviation) del Modelo LC

En el panel **I** se grafican diagramas box-plot para $D_{x,\tau}$ para cada $x = 1, \dots, n_a$, y en el panel **II**, para cada cohorte, $\tau = 98/01, \dots, 06/01$. Inicialmente se observa que hay varianza constante a pesar de que los primeros semestres 1,2,3 presentan mayor dispersión, posiblemente debido a que la deserción es más pronunciada en éstos. En el panel **II** se observa una posible heterocedasticidad, debido a que los semestres del año 2000 en adelante muestran mayor dispersión. Lo anterior puede afectar la calidad y eficiencia de los estimadores generados. En los paneles **III** y **IV** se presentan los residuales por cohorte para los estudiantes que egresan en 10 semestres y los que egresan en más de 10 semestres respectivamente. Haciendo esta distinción se observa que no hay cambios a lo largo de las cohortes entre las observaciones de

estudiantes que salen de la universidad en menos de 10 semestres y los demás, evidentemente en la gráfica se observan unos valores outliers que afectan la distribución de datos en algunos periodos.

Para reforzar la decisión de seguir utilizando el modelo implementado con la librería `gnm` se muestra la tabla **5-1** en la que se comparan los residuales de la estimación por descomposición singular (2-17) vs. la estimación por minimización del estadístico deviance (2-22) equivalente a la máxima verosimilitud en el modelo log-bilineal de Poisson, mediante el valor del estadístico chi-cuadrado Deviance (2-22) y el estadístico chi-cuadrado Pearson definido como:

$$\chi_P^2 = \sum_{x,\tau} \frac{(d_{x,\tau} - \hat{d}_{x,\tau})^2}{\hat{d}_{x,\tau}}. \quad (5-1)$$

Como se puede observar el ajuste mediante máxima verosimilitud con el modelo log-bilineal de Poisson resulta mejor ya que la varianza del error cuadrático es menor bajo ambas medidas (Deviance y Pearson).

	SVD	Max.Ver
Deviance	2090.93	1699.17
Pearson	2188.70	1645.29

Tabla 5-1: Comparación de residuales Deviance y Pearson

5.2. Proyección de mortalidad

La mortalidad se estimó a través de la proyección de la serie de tiempo generada por el parámetro κ_τ . Este ejercicio se realizó ajustando dos modelos, el primero un ARIMA(1,1,0) con parámetro autoregresivo $\hat{\varphi} = -0,3419$ y error estándar 0.1666 determinado a través de la función `auto.arima` de la librería `forecast` de [Hyndman et al., 2014] para R. En segundo lugar se realizó una regresión local Loess con la función `loess()` con un ancho de ventana `span=0.42` (no se utilizó el valor por defecto de 0.75). El cambio en el ancho de ventana corresponde a que luego de varios ensayos se concluyó que 0,42 generaría el mejor ajuste.

Los resultados se observan en la figura **5-5**. Para un pronóstico de 10 semestres entre 2006–02 y 2011–01 la tendencia generada por el modelo ARIMA(1,1,0) es constante, a diferencia de la tendencia generada por el modelo Loess, la cual es decreciente.

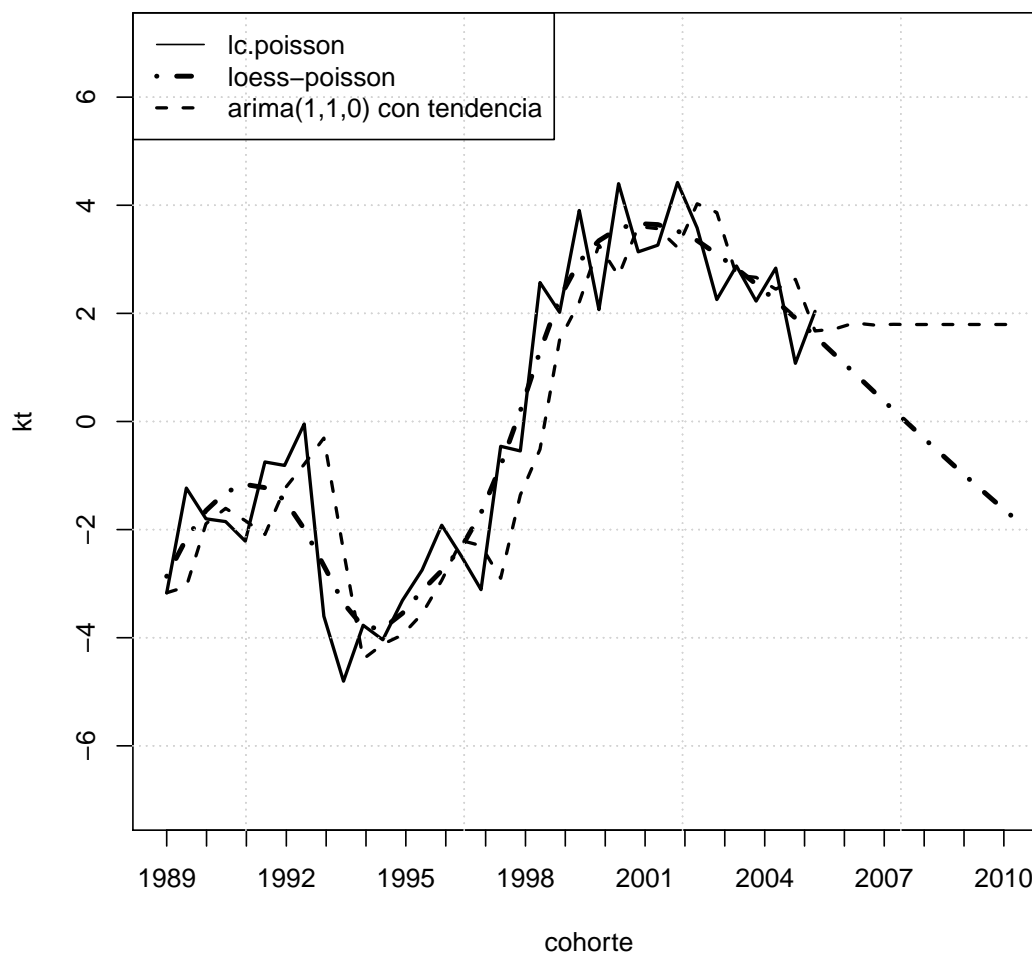


Figura 5-5: Ajuste de κ_t con ARIMA y Loess, y pronóstico a 10 semestres

Dado que al momento del estudio se cuenta con la información hasta el octavo semestre de la cohorte del 2011-01 (κ_{34+10}) se puede establecer cual genera un mejor pronóstico de la mortalidad académica. A partir de esta información se reconstruye la tasa central de mortalidad $m_{x,2011-01}$ y los valores $l_{x+1} = l_x e^{-\mu_{x,2011-01}}$, para los semestres $x = 1, \dots, 17$.

	observada	pron.Loess	pron.Arima
1	1465	1465	1465
2	1124	1291	1269
3	1107	1151	1115
4	1007	1043	1004
5	961	972	942
6	908	924	896
7	878	894	864
8	851	869	836
9		849	807
10		823	758
11		774	663
12		690	530
13		570	388
14		437	276
15		320	198
16		231	144
17		161	104

Tabla 5-2: Semestre 2011-01, pronosticada con Loess y Arima; Fuente de datos observados: Oficina de planeación académica, Universidad Nacional Sede Medellín

Para 2014-02 se conocían los valores de 8 de estos 17 semestres; son: 1465, 1124, 1107, 1007, 961, 908, 878 y 851. Utilizando $l_1 = 1465$ los valores obtenidos a partir de la estimación de $\mu_{x,2011-01}$ por Loess y ARIMA(1,1,0) están en la Tabla 5-2 siguiente, y en la Figura 5-6. En ambas se observa que, excepto los 4 primeros semestres, los pronósticos con Loess sobre-estiman mientras que con ARIMA(1,1,0) se sub-estiman los valores observados en 2014 para los primeros 8 periodos del semestre 2011-01. Mediante una revisión del error cuadrático de la estimación se concluye que el mejor ajuste, y por ende la mejor proyección, es la presentada por el modelo ARIMA(1,1,0).

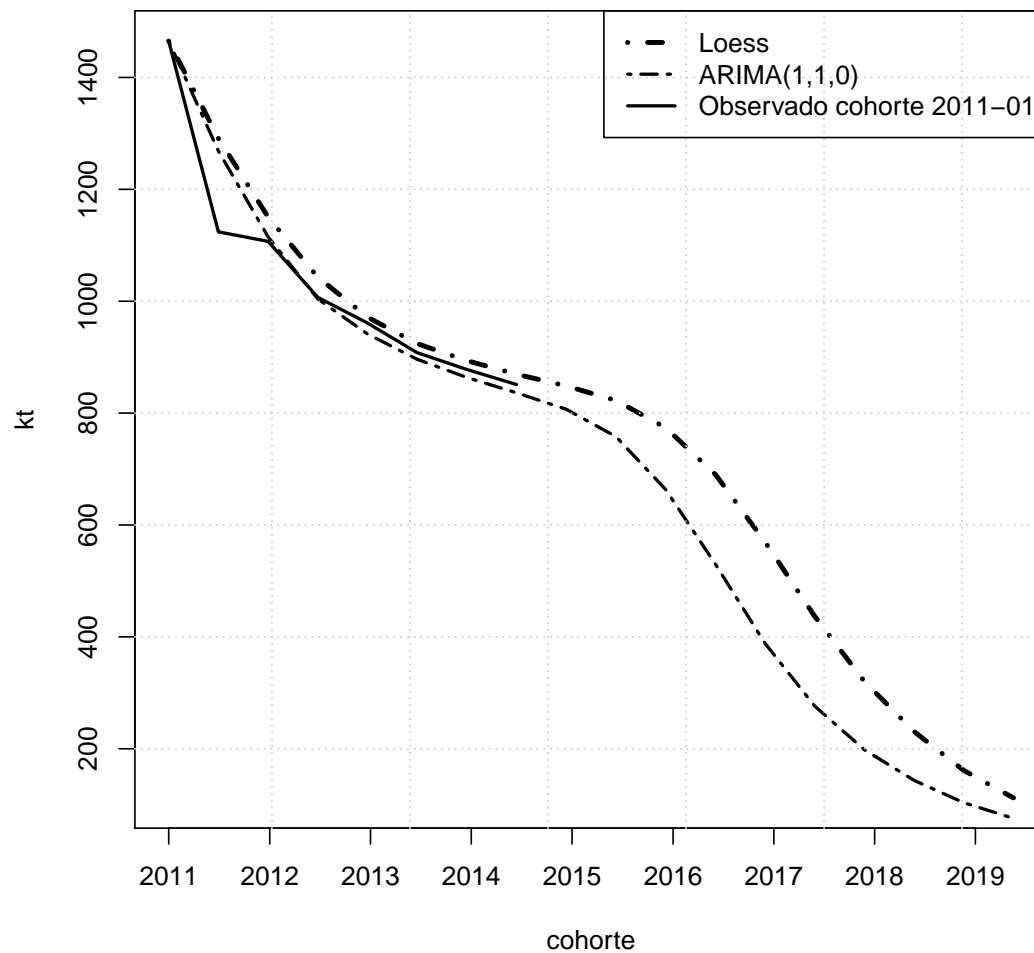


Figura 5-6: Pronósticos de 17 semestres a partir del 2011-01 con modelo LC usando Loess y ARIMA, datos observados para 8 semestres (Fuente: Oficina de Planeación, Universidad Nacional de Colombia Sede Medellín)

6 Conclusiones y recomendaciones

6.1. Conclusiones

- El modelo LC sugiere que la mortalidad en Colombia ha venido decreciendo en las últimas 5 décadas y, a través de los resultados de pronósticos Loess, se puede afirmar que este factor poblacional seguirá disminuyendo a través del tiempo.
- A partir de la información analizada se puede sugerir que la mortalidad masculina es menos cambiante que la mortalidad femenina, especialmente para edades entre los 20 y los 60 años, lo que por ende permite concluir que en el futuro la esperanza de vida para las mujeres será mejor que la de las hombres.
- Es valido afirmar que la implementación de un modelo como el LC para la información demográfica de un país puede sugerir dinámicas poblacionales que permitan revisar las formulaciones que basan los modelos previsionales de fondos de pensiones y seguros. Es evidente que la disminución en la tasa de la mortalidad llevará a un costo más alto de rentas vitalicias, lo que pone en riesgo el modelo previsional dado las cifras en las que está basado por ley.
- Con base en los resultados obtenidos se puede concluir que la evolución de la población universitaria muestra un incremento en la deserción en los primeros semestres, 1 a 4, a partir de 2001, y al mismo tiempo, una disminución del tiempo requerido para graduarse en 10 semestres.
- La utilidad de esta metodología puede verse desde la planeación académica ya que sería posible analizar indicadores tales como número de estudiantes por docente, deserción estudiantil, permanencia promedio en la universidad, entre otros.
- La metodología original de LC con base en la descomposición en valores singulares no parece fácil de implementar en modelos alternos de LC como los propuestos en [Cairns et al., 2009], por lo que la implementación en R de la estimación con base en modelos no lineales generalizados resulta de suma utilidad y permite la extensión de otros modelos, por ejemplo, incorporando decrementos múltiples (riesgos competitivos)

6.2. Recomendaciones

- Bajo el supuesto de que las empresas de seguros y pensiones están concentradas regionalmente, se sugiere realizar un estudio de la dinámica poblacional que diferencie la tasa de mortalidad por regiones. Lo anterior podrá sugerir un riesgo diferencial por region y por ende a costos previsionales más altos o más bajos para diferentes participantes en sector.
- De manera similar, un estudio de mortalidad académica por sede podrá sugerir en qué lugares se están evidenciando con mayor fuerza las políticas implementadas en términos académicos y de regulación para determinar qué condiciones particulares afectan a las diferentes poblaciones. Así mismo se pueden generar estudios adicionales que incorporen variables como estado civil, ocupación laboral, estrato, género, etc. de forma que identifiquen factores externos que afecten la mortalidad académica.

Bibliografía

- [Betzuen, 2010] Betzuen, A. (2010). Un análisis sobre las posibilidades de predicción de la mortalidad futura aplicando el modelo lee-carter. *Anales*.
- [Bowers et al., 1986] Bowers, N., Gerber, H., Hickman, J., Jones, D., & Nesbitt, C. J. (1986). *Actuarial Mathematics*. Ithasca, IL: The Society of Actuaries.
- [Brouhns & Denuit, 2002] Brouhns, N. & Denuit, M. (2002). Risque de longévité et rentes viagères II. Tables de mortalité prospectives pour la population belge. *Belgian Actuarial Bulletin*, 2(1), 50–63.
- [Brouhns et al., 2002] Brouhns, N., Denuit, M., & Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31, 373–393.
- [Butt & Haberman, 2009] Butt, Z. & Haberman, S. (2009). Ilc: A Collection of R Functions for Fitting a Class of Lee-Carter Mortality Models using Iterative Fitting Algorithms. *Actuarial Research Paper No. 190, Faculty of Actuarial Science and Insurance, Cass Business School*, (pp. 1–37).
- [Cairns et al., 2009] Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., & Balevich, I. (2009). A Quantitative Comparison of Stochastic Mortality Models Using Data from England and Wales and the United States. *North American Actuarial Journal*, 13(1), 1–35.
- [Currie, 2014] Currie, I. (2014). On fitting generalized linear and nonlinear models of mortality. *Scandinavian Actuarial Journal*, 14, 1–28.
- [Davidian, 2009] Davidian, M. (2009). *Nonlinear models for univariate and multivariate response*. North Carolina State University: Lecture Notes. Department of Statistics.
- [Eckart & Young, 1936] Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211–218.
- [Gómez, 1998] Gómez, C. L. M. (1998). Colombia. tablas abreviadas de mortalidad por sexo para fechas censales y estimaciones quinquenales 1995 - 2025. *Estudios Censales*.

- [Gompertz, 1825] Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society*, 115:513-585.
- [Helligman & Pollard, 1980] Helligman, L. & Pollard, J. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107: 49-75.
- [Hyndman et al., 2014] Hyndman, R., Athanasopoulos, G., Razbash, S., Schmidt, D., Zhou, Z., Khan, Y., Bergmeir, C., & Wang, E. (2014). *forecast: Forecasting functions for time series and linear models*. R package version 5.5.
- [Hyndman, 2015] Hyndman, R. J. (2015). Demography: Forecasting mortality, fertility, migration and population data. <http://robjhyndman.com/software/demography/>.
- [Koissi & Shapiro, 2008] Koissi, M.-C. & Shapiro, A. (2008). The Lee-Carter model under the condition of variables age-specific parameters. *43rd Actuarial Research Conference, Regina, Canada*.
- [Lee & Carter, 1992] Lee, R. & Carter, L. R. (1992). Modelling and Forecasting U.S.Mortality. *Journal of the American Statistical Association*, 87, 659–671.
- [Miller, 2011] Miller, T. (2011). Estimación de tendencias de la mortalidad en la niñez: Modelos de regresión. celade, división de población de cepal. http://www.cepal.org/celade/noticias/paginas/3/45123/celade_mort_ninez.pdf.
- [Mundial, 1994] Mundial, B. (1994). *Envejecimiento sin crisis. Políticas para la protección de los ancianos y la promoción del crecimiento*. Washington DC.: Oxford University Press.
- [Renshaw & Haberman, 2009] Renshaw, A. E. & Haberman, S. (2009). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance Mathematics and Economics*.
- [Shores, 2004] Shores, T. (2004). *Applied Linear Algebra and Matrix Analysis*. Heidelberg: Springer Verlag.
- [Turner & Firth, 2015] Turner, H. & Firth, D. (2015). Generalized nonlinear models in r: An overview of the gnm package. v1.0-8. <http://CRAN.R-project.org/package=gnm>.
- [Wilmoth, 1993] Wilmoth, J. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. *Technical Report. Department of Demography, University of California, Berkeley*.